ACL 2024

# 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)

## Proceedings of the Conference

August 11-16, 2024

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

## Diamond

LG AI Research

NEWSBREAK

Meta AI

Google DeepMind

## Platinum

Megagon Labs

Baidu 百度

SCB 10X

SONY

Alibaba Cloud × TONGYI

amazon | science

ByteDance

IBM

Meituan

ORACLE

## Gold

## Silver

## Bronze

## Diversity Inclusion Ally

## Local Sponsors

NECTEC
a member of NSTDA

©2024 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 979-8-89176-094-3

# Message from the General Chair

Originally named the Association for Machine Translation and Computational Linguistics (AMTCL), the Association for Computational Linguistics was founded in 1962 and renamed the ACL in 1968.
The ACL is run by some 20 volunteers overseeing the administration of the Association (organising elections, deciding on new actions, adapting to the fast changing trends of our fields), the publication of two journals (Computational Linguistics and the Transaction of the ACL) and the organisation of the ACL, EACL, NAACL and AACL conferences. The ACL executive board is regularly renewed based on elections from the membership. To ensure continuity, some of the volunteers (e.g., secretary, treasurer and anthology editor) serve longer term mandates and a professional Business Manager gives much needed support to the daily management of the Association.
While ACL runs multiple conferences and workshops each year, the ACL conference is the flagship conference of the Association and like the field in general, it has seen drastic changes over the years both in terms of approaches (from symbolic to statistical to neural) and in terms of popularity (from a few academics to a large population of several thousands academics and industrials with widespread geographical coverage).
It is a testimony to the strength of the Association that the ACL meeting has been held annually since 1962. Also remarkable is that, thanks to Steven Bird's initiative, the proceedings are available online in the ACL anthology since ACL 1979 (which hosted a whopping 23 papers!).
In this note, I would like to welcome all participants to ACL 2024, the 62$^{nd}$ Annual Meeting of the Association for Computational Linguistics (held in Bangkok, Thailand, August 11-16, 2024), and to express my gratitude to another large set of volunteers who made ACL 2024 possible.
First and foremost, I would like to thank the three Program Chairs, Lun-Wei Ku, Andre Martins and Vivek Srikumar who oversaw the reviewing process, selected the keynote speakers and created the conference program. 2024 was the first year where all *CL conferences moved entirely to the ACL Rolling Review model. This meant a novel process working in tight interaction with the ARR team while handling a huge number of submissions. Thanks to Lun-Wei, Andre and Vivek, to the ARR Editors in Chief (Mausam, Viviane Moreira, Vincent Ng, Lilja Øvrelid, Thamar Solorio, Jun Suzuki), to the Senior Area Chairs, Area Chairs, reviewers, and to the Best Paper Committee, who worked together to select the ACL 2024 program.
The workshop and the tutorial program was created by the Workshop (Eunsol Choi, Xipeng Qiu) and the Tutorial (Luis Chiruzzo, Hung-yi Lee, Leonardo Ribeiro) Chairs, who collaborated with EACL, NAACL and EMNLP 2024 to select 32 workshops and 6 tutorials that cover both technical and societal areas of Natural Language Processing.
The program also includes demonstrations selected by the Demonstration Chairs (Yixin Cao, Yang Feng, Deyi Xiong), as well as the traditional Student Research Workshop, which was put together by the SRW Chairs (Eve Fleisig, Xiyan Fu), with the guidance and support of the faculty advisors (Ekapol Chuangsuwanich, Yuval Pinter). Thank you all!
Preparing the proceedings is another large, time consuming task. Thanks to the Publication Chairs (Miruna Clinciu, Zhiyu Zoey Chen, Chen Liang, Bing Liu) for coordinating the preparation of all proceedings, including the main conference proceedings, findings, demonstration, SRW and workshop proceedings.
The Website Chairs (Yun-Nung Chen, Vipas Sutantayawalee) created the website and were quick in responding to our queries for updates; thank you Vipas and Yun-Nun. Thanks also to the Publicity and Social Media Chairs (Yuki Arase, Dimitra Gkatzia, Jing Jiang) who communicated and publicized the conference through various social media channels, enhancing the visibility and reach of the conference. Thanks also to the Handbook Chairs (Loic Barrault, Pierre Colombo) for creating the conference handbook that will guide you through the conference program.
Making ACL accessible to a wider community was the task of the Diversity and Inclusion Chairs (Aparna Garimella, Lin Gui, Jing Li, Steven Wilson). Thank you all for helping in fostering a diverse and inclusive environment.

Thanks to the Ethics Chairs (Aurélie Névéol, Alice Oh), who checked papers flagged with ethics issues. They had to process many more papers than expected but handled the overflow brilliantly.
The Technical Open Review Chairs (Thiago Castro-Ferreira, Taro Watanabe) helped out with Open Review related requests and the Virtual Infrastructure Chairs (Gaël Guibon, Gözde Gül, Rachada Kongkrachantra), made various enhancements to the virtual platform to ensure an engaging conference experience. Thank you!
The conference is also a successful recipient for thousands of emails. Many thanks go to the Internal Communication Chairs (Claudia Borg, Yannick Parmentier, Valentina Pyatkin), for their efficient and much needed processing of the multiple emails sent to ACL 2024.
The registration fees would be considerably higher without our sponsors generous contributions. Sincere thanks to them and to Chris Callison-Burch, the ACL sponsorship Director who, together with the Sponsorship Chairs (Lluis Marquez, Kobbrit Viriyayudhakorn) succeeded in securing the sponsorships that are crucial in helping keep registration fees down.
Sol Rosenberg and his team provide the Underline virtual platform - thank for their support in collaborating with us to meet our needs and accommodate ACL feature requests.
The local team (Thepchai Supnithi, Prachya Bookwan, Thanaruk Theeramunkong) did a wonderful job locating the venue, providing help with local and visa information, booking hotels for the participants and organizing a social program.
During the conference, student volunteers help make the conference run smoothly. Many thanks to them and to the Student Volunteer Chairs (Hao Fei, Margot Mieskes, Liangming Pan) who reviewed applications, selected the student volunteers, and assigned them their tasks.
Much of the know-how for the various chair positions is based on insights gleaned from earlier events. Thanks to previous ACL conference Chairs for sharing their experience, and to the ACL Exec for their support.
Special thanks go to Jennifer Rachford, ACL Business Manager, whose remarkable ability to juggle multiple conferences and interact with overloaded scientists make the impossible possible. Her knowledge of the ACL conferences and of their multiple aspects is essential for the success of our conferences.
Finally, let me thank you all, authors, reviewers, presenters, workshop organizers and participants of the conference. Thank you for choosing to be part of ACL 2024, I wish you a very enjoyable conference!

ACL 2024 General Chair
Claire Gardent
CNRS, France

# Message from the Program Chairs

Welcome to the 62nd Annual Meeting of the Association for Computational Linguistics! ACL 2024 will feature a hybrid format, allowing attendees to join us in person in Bangkok, Thailand, or to participate remotely from anywhere in the world. We are pleased to be hosting the conference in Bangkok, which was the original planned venue for ACL 2021 before the COVID-19 pandemic forced a change.
Organizing ACL 2024 has been a team effort involving thousands of people. We would like to thank the support and contributions of the following people:

- The General Chair, Claire Gardent;
- The ARR Editors-in-Chief of the Feb 2025 cycle (Viviane Moreira, Jun Suzuki) and the entire team (Lilja Øvrelid, Mausam, Thamar Solorio, Vincent Ng, Jonathan Kummerfield, Sudipta Kar);
- The OpenReview team;
- The 72 Senior Area Chairs;
- The 718 Area Chairs and the 4209 reviewers;
- The awards committee chairs, Hal Daumé, Raquel Fernández and Yuji Matsumoto, and the 20 awards committee members;
- The ethics committee led by Alice Oh and Aurélie Névéol, along with Malihe Alikhani and Vinodkumar Prabhakaran from ARR;
- The website chairs, Yun-Nung (Vivian) Chen and Vipas Sutantayawalee;
- The publication chairs, Miruna Clinciu, Bing Liu, Zhiyu Zoey Chen and Chen Liang;
- The handbook chairs Pierre Colombo and Loic Barrault
- The local organization chairs, Thepchai Supnithi, Prachya Bookwan, Thanaruk Theeramunkong, and their team;
- The publicity and social media chairs, Yuki Arase, Jing Jiang, and Dimitra Gkatzia;
- The student volunteer chairs, Margot Mieskes, Hao Fei and Liangming Pan;
- The ACL Anthology Director, Matt Post, and his team;
- The TACL editors-in-chief (Asli Celikyilmaz, Roi Reichart, Dilek Hakkani Tur) and CL Editor-in-Chief Wei Lu for coordinating TACL and CL presentations with us;
- The NAACL 2024 Program Chairs (Kevin Duh, Helena Gomez, and Steven Bethard) and the ACL 2023 Program Chairs (Anna Rogers, Jordan Boyd-Graber and Naoki Okazaki);
- Damira Mršić and Underline Team;
- Jennifer Rachford and entire conference support staff;
- All the authors of papers submitted for review and committed to the conference.

**Review process** All ACL 2024 submissions were channeled through a two-stage review process: Submissions were first sent to ACL Rolling Review (ARR) for reviews (by reviewers) and meta-reviews (by area chairs). Then, authors could choose to commit their reviewed papers to ACL via a separate ACL 2024 commitment site for recommendations by senior area chairs and final acceptance decisions by the program chairs. In this, ACL 2024 follows EACL 2024 and NAACL 2024.
We worked closely with the ARR team, especially the February 2024 Editors-in-Chief, and served as guest Editors-in-Chief for this cycle. We helped recruit new reviewers and area chairs to ARR, resulting in 4209 reviewers and 718 ACs in the 2024 February ARR cycle to which most ACL 2024 papers were submitted. The 72 senior area chairs recruited by ACL helped oversee the review and meta-review process during this phase. Overall, the ARR process went mostly smoothly, successfully delivering at least three reviews and a meta-review for all papers submitted.
For the ACL commitment part of the process, the senior area chairs made acceptance recommendations for 2931 committed papers based on the papers, reviews, and meta-reviews, and program chairs finalized the recommendations into acceptance decisions.

**Acceptance rate** The acceptance rate calculation follows precedent set by previous conferences that also go through ARR, e.g. EACL 2024, NAACL 2024. The calculation takes into account the multi-stage process of ARR where a paper may get revised in ARR and then later committed to the conference. The denominator includes:

- Papers in the ARR February 2024 cycle that selected ACL as a preferred venue.
- Papers in the ARR February 2024 cycle that did not select any conference as a preferred venue.
- Papers in the ARR February 2024 cycle that selected another conference, but then committed to ACL 2024.
- Papers in the ARR cycles before February 2024 that committed to ACL 2024.

In total, we had 4,835 submissions in the ARR February 2024 cycle. Among these, 276 were withdrawn before reviews were released and 169 were desk rejected. Among the remaining, 4244 had either an unspecified venue or included ACL as a desired venue. Among the submissions that selected other venues, three papers were committed to ACL. Finally, an additional 160 papers from other cycles were committed to ACL. In total, the denominator for the acceptance rate calculation is 4244 + 3 + 160= 4407. Among these, 2931 were committed to ACL.
Among the committed papers, 940 were accepted to the Main Conference. The acceptance rate for Main Conference papers is therefore 940 / 4407 = 21.3%. A further 975 papers were accepted to Findings of ACL, representing solid work that is has been judged worthy of publication with sufficient substance, quality and novelty. The acceptance rate for Findings of ACL is 975 / 4407 = 22.1%.

**Special Theme: Open science, open data, and open models for reproducible NLP research** The rise of large language models as a general purpose tool for NLP has opened up exciting possibilities for NLP. But their widespread adoption via closed APIs has also raised concerns about transparency and reproducibility. When we do not have access to information about how these models were trained or the data they learned from, it becomes challenging to build upon existing research and compare new approaches fairly. This lack of openness poses a risk to progress in our field. With this perspective in mind, for ACL 2024, we invited submissions to a special theme titled "Open science, open data, and open models for reproducible NLP research".
With this theme track, we sought a discussion on increased transparency in the field by promoting the use of open models and open-source initiatives in NLP as an alternative to closed approaches. We encouraged contributions related to the release of high quality datasets, novel ideas for evaluation, non-trivial algorithm and toolbox implementations, and models which are properly documented (e.g. via model cards).

We received 55 submissions to the theme track during the review phase. Among these, 22 papers were accepted to the main conference and a further 16 to Findings of ACL.

**Paper Awards** ACL 2024 implemented the updated ACL award policy that seeks to expand the pool of work recognized as outstanding. In total 102 papers were nominated by the reviewers, area chairs and senior area chairs for consideration. The Awards Committee assessed these papers to select the best papers (featuring $\leq 0.25\%$ of accepted papers), outstanding papers (featuring $\leq 2.5\%$ of accepted papers), and special awards for social impact and best resource. Separately, the senior area chairs of individual tracks also selected papers in their track for the area chair awards. Finally, the program chairs selected one paper from the papers submitted to the conference theme track as the best theme paper.
The recent change in ACL policy allows papers to be non-anonymous during the review process via public preprints. To recognize submissions that remained anonymous, we followed the policy recommendation to have separate best and outstanding paper awards for submissions that remained anonymous to the public during the whole process.
All the awards will be announced in a dedicated plenary session on the last day of the conference.

**Program composition & presentation modes** Based on feedback from the conference support staff and the Underline team after NAACL 2024, we decided to hold the virtual poster session separately, and *after* the in-person conference ended. The post-conference virtual sessions were scheduled to avoid conflicts with the in-person attendees of the conference. The goal was to encourage all attendees — both virtual and in-person — to join the virtual conference. All in-person papers accepted to the main conference were given a poster slot. In addition, 102 papers were assigned oral presentations. These papers were selected by the program chairs and the decision was motivated by the goal of having a well-rounded program with a diverse set of topics. Additionally, all Findings papers were also assigned a poster presentation in separate Findings posters sessions in the conference.
The ACL program features three exciting keynote speakers: Sunita Sarawagi, Subbarao Kambhampati and Barbara Plank. Further, to celebrate the venue of ACL, the program also including a panel discussion focusing on Southeast Asian languages featuring panelists Ayu Purwarianti, William Tjhi and Sarana Nutanong.
The ACL program also includes 6 papers accepted by the Computational Linguistics journal and 31 papers accepted by the Transactions of ACL (TACL). All journal papers whose authors were attending the conference in person were given oral presentation slots and thematically distributed in the conference in appropriate sessions. The program is rounded out with dedicated sessions during the main conference for the demonstrations track and student research workshop.
We hope that you will enjoy this year's program and conference!

Lun-Wei Ku (Academia Sinica, Taiwan)
André F. T. Martins (Instituto Superior Técnico, Instituto de Telecomunicações, Unbabel, Portugal)
Vivek Srikumar (University of Utah, USA)
ACL 2024 Program Committee Co-Chairs

# Organizing Committee

**General Chair**

Claire Gardent, CNRS and Université de Lorraine

**Program Chairs**

Lun-Wei Ku, Academia Sinica
Andre Martins, Instituto Superior Técnico / Instituto de Telecomunicações / Unbabel
Vivek Srikumar, University of Utah

**Local Organization**

Thepchai Supnithi, NECTEC and AIAT
Prachya Bookwan, NECTEC and AIAT
Thanaruk Theeramunkong, SIIT and AIAT

**Workshop Chairs**

Xipeng Qiu, Fudan University
Eunsol Choi, The University of Texas at Austin

**Tutorial Chairs**

Luis Chiruzzo, Universidad de la República
Hung-yi Lee, National Taiwan University
Leonardo Ribeiro, Amazon Alexa Seattle

**Demonstration Chairs**

Yixin Cao, Singapore Management University
Yang Feng, Chinese Academy of Science
Deyi Xiong, Tianjin University

**Student Research**

Xiyan Fu, Heidelberg University
Eve Fleisig, UC Berkeley

**Student Research: Faculty Advisor**

Ekapol Chuangsuwanich, Chulalongkorn University
Yuval Pinter, Ben Gurion University

**Publicity and Social Media Chairs**

Yuki Arase, Osaka University
Jing Jiang, Singapore Management University

Dimitra Gkatzia, Napier University

**Publication Chairs**

Miruna Clinciu, University of Edinburgh
Bing Liu, Meta AI
Zhiyu Zoey Chen, University of Texas at Dallas
Chen Liang, Google DeepMind

**Handbook Chairs**

Pierre Colombo, Université Paris Saclay
Loic Barrault, Meta AI

**Technical Open Review**

Taro Watanabe, Nara Institute of Science and Technology (NAIST)
Thiago Castro, Federal University of Minas Gerais

**Diversity and Inclusion**

Jing Li, Hong Kong Polytechnic University
Aparna Garimella, Adobe Research
Steven Wilson, Oakland University
Lin Gui, King's College London

**Ethic Committee**

Alice Oh, KAIST
Aurélie Névéol, CNRS and Université Paris Saclay

**Internal Communications**

Claudia Borg, University of Malta
Valentina Pyatkin, Allen Institute for AI
Yannick Parmentier, Université de Lorraine

**Student Volunteer**

Margot Mieskes, University of Applied Science, Darmstadt
Hao Fei, National University of Singapore
Liangming Pan, University of California, Santa Barbara

**Virtual Infrastructure**

Gözde Gül, Koç University
Gael Guibon, University of Lorraine
Rachada Kongkrachantra, Thammasat University

**Website and Conference App**

Yun-Nung (Vivian) Chen, National Taiwan University
Vipas Sutantayawalee, Artificial Intelligence Entrepreneur Association of Thailand

**Sponsorship Chairs**

Lluis Marquez, Amazon
Kobkrit Viriyayudhakorn, iAPP Co., Ltd

# Program Committee

## SRW Program Chairs

Miruna Clinciu
Eve Fleisig
Xiyan Fu

## SRW Reviewers

Piush Aggarwal, Chandan Akiti, Abeer Aldayel, Miguel A. Alonso, Begoa Altuna, Sergi Alvarez-Vidal, Bharat Ram Ambati, Evelin Amorim, Tatiana Anikina, Raghav Arora

Parsa Bagherzadeh, Ahsaas Bajaj, Valerio Basile, Anil Batra, Tim Baumgrtner, Rachel Bawden, Shabnam Behzad, Gabriel Bernier-Colborne, Anne Beyer, Isil Poyraz Bilgin, Shikha Bordia

Meng Cao, Arlene Casey, Silvia Casola, Canyu Chen, William Chen, Hyundong Justin Cho, Sukmin Cho, Juhwan Choi, Arijit Ghosh Chowdhury, Ekapol Chuangsuwanich, Rodolfo Corona, Amanda Cercas Curry

Aryia Dattamajumdar, Brian Davis, Alok Debnath, Ayse Ayyuce Demirbas, Nesat Dereli, Chris Develder, Noshin Tabassum Dina, Bonaventure F. P. Dossou, Zi-Yi Dou, Ritam Dutt, Aditi Dutta

Micha Elsner

Neele Falk, Antnio Farinhas, Fauzan Farooqui, Eve Fleisig, Francesca Franzon, Dayne Freitag, Xiyan Fu

Diana Galvan, Shubham Garg, Aina Gar Soler, Ratul Ghosh, Kristina Gligoric, Anmol Goel, Satya Kesav Gundabathula, Ankita Gupta

Ivan Habernal, Youssef Al Hariri, Yifan Hou, Phu Mon Htut, Tiancheng Hu, Zhongzhan Huang

Robert L. Logan IV

Glorianna Jagfeld, Labiba Jahan, Thanmay Jayakumar, Jyoti Jha, Kenan Jiang, Kristen Johnson

Jad Kabbara, Patrick Kahardipraja, Mihir Kale, Amr Keleg, Yova Kementchedjhieva, Mohammad Aflah Khan, Harshita Khandelwal, Anirudh Khatry, Elena Kochkina, Prashant Kodali, Mamoru Komachi, Fajri Koto, Pranoy Kovuri, Devang Kulshreshtha, Kemal Kurniawan, Mascha Kurpicz-Briki

Alexandra Lavrentovich, Amirhossein Layegh, Messi H.j. Lee, Yoonsang Lee, Jia Li, Manling Li, Sha Li, Siyan Li, Baohao Liao, Jasy Suet Yan Liew, Zheng Wei Lim, Peerat Limkonchotiwat, Lucy H. Lin, Tatiana Litvinova, Wei Liu, Li Lucy, Bohan Lyu

Bolei Ma, Valentin Malykh, Lara J. Martin, Sandeep Mathias, Shubhanshu Mishra, Sai Ajay Modukuri

Hongbin Na, Murali Nandan Nagarapu, Sachin Narayan Nagargoje, Masaaki Nagata, Pakawat

Nakwijit, Kota Shamanth Ramanath Nayak, Nihal V. Nayak, Mariana Neves, Kiet Van Nguyen, Ratna Nirupama

Naoki Otani, Kun Ouyang

Maria Leonor Pacheco, Jiayi Pan, Artemis Panagopoulou, Sara Papi, Tanmay Parekh, Ajay Patel, Naiara Perez, Gabriele Pergola, Ben Peters, Shubham Milind Phal, Van-Thuy Phi, Dina Pisarevskaya, Dongqi Pu, Rajkumar Pujari

Yusu Qian, Yusu Qian

Sunny Rai, Surangika Ranathunga, Pretam Ray, Frederick Riemenschneider, Shruti Rijhwani, Hossein Rouhizadeh, Markian Rybchuk, Sangwon Ryu

Farig Sadeque, Philipp Sadler, Oscar Sainz, Nicholas Sanders, Sashank Santhanam, Abulhair Saparov, Ryohei Sasano, Tatjana Scheffler, Fabian Schmidt, Sebastian Schuster, Indira Sen, Sheikh Shafayat, Aditya Shah, Roshan Sharma, Sina Sheikholeslami, Kejian Shi, Manish Shrivastava, Chenglei Si, Pushpdeep Singh, Kevin Small, Cesare Spinoso-Di Piano, Richard Sproat, Marija Stanojevic, Mihai Surdeanu, Dage Srg, Dage Srg

Shabnam Tafreshi, Zeerak Talat, Alberto Testoni, Nicholas Tomlin, Son Quoc Tran

Can Udomcharoenchaikit

Sowmya Vajjala, Andrea Varga, Francielle Vargas, Deeksha Varshney, David Vilar

Hao Wang, Mingyang Wang, Bonnie Webber, Alexander Wettig, Chenglin Wu, Xinyi Wu

Qi Xin, Rui Xing, Fuzhao Xue

Dixi Yao, Michael Miller Yoder

Olga Zamaraeva, Zenan Zhai, Chenhao Zhang, Haoran Zhang, Mike Zhang, Yangjun Zhang, Yao Zhang, Yiqun Zhang, Zhehao Zhang, Siheng Zhao, Wenjie Zhao, Xin Zhao, Yu Zhao, Xinyi Zheng, Shanshan Zhong, Naitian Zhou, Rongxin Zhu, Elena Zotova

Omri Abend, Hebrew University of Jerusalem
Fernando Alva-Manchego, Cardiff University
Maria Antoniak
Marianna Apidianaki, University of Pennsylvania, University of Pennsylvania
Esma Balkir, Wisetech Global
Maria Barrett, IT University of Copenhagen
Tilman Beck
Terra Blevins, University of Washington
Su Lin Blodgett, Microsoft
Elia Bruni, Universitt Osnabrck
Agostina Calabrese, University of Edinburgh
Ronald Cardenas
Asli Celikyilmaz, FAIR

Ming-Wei Chang, Google
Yen-Chun Chen, Microsoft
Pinzhen Chen, University of Edinburgh
Mingda Chen, Meta AI
Danqi Chen, Department of Computer Science, Princeton University
Christos Christodoulopoulos, Amazon
Alessandra Teresa Cignarella, aequa-tech srl
Volkan Cirik, ASAPP
Xiang Dai, CSIRO
Marco Del Tredici, Amazon
David Demeter
Laura Dietz, University of New Hampshire
Steffen Eger, Universitt Mannheim
Yanai Elazar, Allen Institute for Artificial Intelligence and Department of Computer Science
Amir Feder, Columbia University and Google
Cesar Ferri
David Gaddy, Google
Kuzman Ganchev, Google and School of Engineering and Applied Science
Tianyu Gao
Marcos Garcia, Universidade de Santiago de Compostela
Iker Garca-Ferrero
Jonas Groschwitz
Vivek Gupta, University of Pennsylvania, United States
Barry Haddow, University of Edinburgh
Andreas Hanselowski
Junxian He, Hong Kong University of Science and Technology
Yulan He, King's College London, University of London
Dieuwke Hupkes, Facebook
Mimansa Jaiswal, Norm AI
Zhijing Jin
Tomoyuki Kajiwara, Ehime University
Ehsan Kamalloo, ServiceNow Inc
Sudipta Kar, Amazon
Alina Karakanta, Leiden University
Shima Khanehzar, CSIRO
Daniel Khashabi, Johns Hopkins University
Gunhee Kim, Seoul National University
Mandy Barrett Korpusik, Loyola Marymount University
Julia Kreutzer, Cohere for AI
Kalpesh Krishna, Google
Ilia Kuznetsov, TU Darmstadt
Gorka Labaka, Universidad del Pas Vasco
Gabriella Lapesa, GESIS  Leibniz Institute for the Social Sciences and Heinrich-Heine University Dsseldorf
Ji-Ung Lee, TU Darmstadt
Lei Li, University of Hong Kong
Tianyi Li
None

## Keynote

# Does In-Context-Learning Offer the Best Tradeoff in Accuracy, Robustness, and Efficiency for Model Adaptation?

**Sunita Sarawagi**
Indian Institute of Technology Bombay, India

**08/12/2024** – Time: **09:30 - 10:30** – Room: **Convention Center B1**

**Abstract:** Adapting a model trained on vast amounts of data to new tasks with limited labeled data has long been a challenging problem, and over the years, a diverse range of techniques have been explored. Effective model adaptation requires achieving high accuracy through task-specific specialization without forgetting previous learnings, robustly handling the high variance from limited task-relevant supervision, and doing so efficiently with minimal compute and memory overheads. Recently, large language models (LLMs) have demonstrated remarkable ease of adaptation to new tasks with just a few examples provided in context, without any explicit training for such a capability. Puzzled by this apparent success, many researchers have sought to explain why in-context learning (ICL) works, but we still have only an incomplete understanding. In this talk, we examine this emerging phenomenon and assess its potential to meet our longstanding model adaptation goals in terms of accuracy, robustness, and efficiency.

**Bio:** Sunita Sarawagi researches in the fields of databases, machine learning, and applied NLP. She got her PhD in databases from the University of California at Berkeley and a bachelors degree from IIT Kharagpur. She has also worked at Google Research, CMU, and IBM Almaden Research Center. She is an ACM fellow, was awarded the Infosys Prize in 2019 for Engineering and Computer Science, and the distinguished Alumnus award from IIT Kharagpur. She has several publications in database, machine learning, and NLP conferences including notable paper awards at ACM SIGMOD, ICDM, and NeurIPS conferences.

## Keynote

# Can LLMs Reason and Plan?

**Subbarao Kambhampati**
Arizona State University, USA

**08/13/2024** – Time: **09:00 - 10:00** – Room: **Convention Center B1**

**Abstract:** Large Language Models (LLMs) are on track to reverse what seemed like an inexorable shift of AI from explicit to tacit knowledge tasks. Trained as they are on everything ever written on the web, LLMs exhibit "approximate omniscience"–they can provide answers to all sorts of queries, but with nary a guarantee. This could herald a new era for knowledge-based AI systems–with LLMs taking the role of (blowhard?) experts. But first, we have to stop confusing the impressive style/form of the generated knowledge for correct/factual content, and resist the temptation to ascribe reasoning, planning, self-critiquing etc. powers to approximate retrieval by these n-gram models on steroids. We have to focus instead on LLM-Modulo techniques that complement the unfettered idea generation of LLMs with careful vetting by model-based verifiers (the models underlying which themselves can be teased out from LLMs in semi-automated fashion). In this talk, I will reify this vision and attendant caveats in the context of our ongoing work on understanding the role of LLMs in planning tasks.

**Bio:** Subbarao Kambhampati is a professor of computer science at Arizona State University. Kambhampati studies fundamental problems in planning and decision making, motivated in particular by the challenges of human-aware AI systems. He is a fellow of Association for the Advancement of Artificial Intelligence, American Association for the Advancement of Science, and Association for Computing machinery. He served as the president of the Association for the Advancement of Artificial Intelligence, a trustee of the International Joint Conference on Artificial Intelligence, the chair of AAAS Section T (Information, Communication and Computation), and a founding board member of Partnership on AI. Kambhampati's research as well as his views on the progress and societal impacts of AI have been featured in multiple national and international media outlets. He can be followed on Twitter @rao2z.

## Keynote

# Are LLMs Narrowing Our Horizon? Let's Embrace Variation in NLP!

**Barbara Plank**
Ludwig Maximilian University of Munich, Germany

**08/14/2024** – Time: **09:00 - 10:00** – Room: **Convention Center B1**

**Abstract:** NLP research has made significant progress, and our community's achievements are becoming deeply integrated in society. The recent paradigm shift due to rapid advances in Large Language Models (LLMs) offers immense potential, but also led NLP to become more homogeneous. In this talk, I will argue for the importance of embracing variation in research, which will lead to more innovation, and in turn, trust. I will give an overview of current challenges and show how they led to the loss of trust in our models. To counter this, I propose to embrace variation in three key areas: inputs to models, outputs of models and research itself. Embracing variation holistically will be crucial to move our field towards more trustworthy human-facing NLP.

**Bio:** Barbara Plank is Professor and co-director of the Center for Information and Language Processing at LMU Munich. She holds the Chair for AI and Computational Linguistics at LMU Munich and is an affiliated Professor at the Computer Science department at the IT University of Copenhagen. Her MaiNLP research lab (Munich AI and NLP lab, pronounced "my NLP") focuses on robust machine learning for Natural Language Processing with an emphasis on human-inspired and data-centric approaches. Her research has been funded by distinguished grants, including an Amazon Research Award (2018), the Danish Research Council (Sapere Aude Research Leader Grant, 2020-2024), and the European Research Council (ERC Consolidator Grant, 2022-2027). Barbara is a Scholar of ELLIS (the European Laboratory for Learning and Intelligent Systems) and regularly serves on international committees, including the Association for Computational Linguistics (ACL), the European Chapter of the ACL, and the Northern European Association for Language Technology (NEALT).

# Table of Contents

*Quantized Side Tuning: Fast and Memory-Efficient Tuning of Quantized Large Language Models*
Zhengxin Zhang, Dan Zhao, Xupeng Miao, Gabriele Oliaro, Zhihao Zhang, Qing Li, Yong Jiang and Zhihao Jia . . . . . 1

*Unsupervised Multimodal Clustering for Semantics Discovery in Multimodal Utterances*
Hanlei Zhang, Hua Xu, Fei Long, Xin Wang and Kai Gao . . . . . 18

*MAGE: Machine-generated Text Detection in the Wild*
Yafu Li, Qintong Li, Leyang Cui, Wei Bi, Zhilin Wang, Longyue Wang, Linyi Yang, Shuming Shi and Yue Zhang . . . . . 36

*PrivLM-Bench: A Multi-level Privacy Evaluation Benchmark for Language Models*
Haoran Li, Dadi Guo, Donghao Li, Wei Fan, Qi Hu, Xin Liu, Chunkit Chan, Duanyi Yao, Yuan Yao and Yangqiu Song . . . . . 54

*GenTranslate: Large Language Models are Generative Multilingual Speech and Machine Translators*
Yuchen Hu, Chen Chen, Chao-Han Huck Yang, Ruizhe Li, Dong Zhang, Zhehuai Chen and EngSiong Chng . . . . . 74

*Exploring Chain-of-Thought for Multi-modal Metaphor Detection*
Yanzhi Xu, Yueying Hua, Shichen Li and Zhongqing Wang . . . . . 91

*BitDistiller: Unleashing the Potential of Sub-4-Bit LLMs via Self-Distillation*
DaYou Du, Yijia Zhang, Shijie Cao, Jiaqi Guo, Ting Cao, Xiaowen Chu and Ningyi Xu . . . . . 102

*A Unified Temporal Knowledge Graph Reasoning Model Towards Interpolation and Extrapolation*
Kai Chen, Ye Wang, Yitong Li, Aiping Li, Han Yu and Xin Song . . . . . 117

*Unsupervised Information Refinement Training of Large Language Models for Retrieval-Augmented Generation*
Shicheng Xu, Liang Pang, Mo Yu, Fandong Meng, Huawei Shen, Xueqi Cheng and Jie Zhou 133

*CSCD-NS: a Chinese Spelling Check Dataset for Native Speakers*
Yong Hu, Fandong Meng and Jie Zhou . . . . . 146

*Evaluating Dynamic Topic Models*
Charu Karakkaparambil James, Mayank Nagda, Nooshin Haji Ghassemi, Marius Kloft and Sophie Fellenz . . . . . 160

*How Abilities in Large Language Models are Affected by Supervised Fine-tuning Data Composition*
Guanting Dong, Hongyi Yuan, Keming Lu, Chengpeng Li, Mingfeng Xue, Dayiheng Liu, Wei Wang, Zheng Yuan, Chang Zhou and Jingren Zhou . . . . . 177

*Through the Lens of Split Vote: Exploring Disagreement, Difficulty and Calibration in Legal Case Outcome Classification*
Shanshan Xu, Santosh T.y.s.s, Oana Ichim, Barbara Plank and Matthias Grabmair . . . . . 199

*Inference to the Best Explanation in Large Language Models*
Dhairya Dalal, Marco Valentino, Andre Freitas and Paul Buitelaar . . . . . 217

*A Novel Cartography-Based Curriculum Learning Method Applied on RoNLI: The First Romanian Natural Language Inference Corpus*
Eduard Poesina, Cornelia Caragea and Radu Tudor Ionescu . . . . . 236

*MinPrompt: Graph-based Minimal Prompt Data Augmentation for Few-shot Question Answering*
Xiusi Chen, Jyun-Yu Jiang, Wei-Cheng Chang, Cho-Jui Hsieh, Hsiang-Fu Yu and Wei Wang 254

*SportsMetrics: Blending Text and Numerical Data to Understand Information Fusion in LLMs*
Yebowen Hu, Kaiqiang Song, Sangwoo Cho, Xiaoyang Wang, Hassan Foroosh, Dong Yu and Fei Liu . . . . . 267

*SciMON: Scientific Inspiration Machines Optimized for Novelty*
Qingyun Wang, Doug Downey, Heng Ji and Tom Hope . . . . . 279

*Expedited Training of Visual Conditioned Language Generation via Redundancy Reduction*
Yiren Jian, Tingkai Liu, Yunzhe Tao, Chunhui Zhang, Soroush Vosoughi and Hongxia Yang . 300

*Confidence Under the Hood: An Investigation into the Confidence-Probability Alignment in Large Language Models*
Abhishek Kumar, Robert Morabito, Sanzhar Umbet, Jad Kabbara and Ali Emami . . . . . 315

*Retrieval-Augmented Multilingual Knowledge Editing*
Weixuan Wang, Barry Haddow and Alexandra Birch . . . . . 335

*Picturing Ambiguity: A Visual Twist on the Winograd Schema Challenge*
Brendan Park, Madeline Janecek, Naser Ezzati-Jivan, Yifeng Li and Ali Emami . . . . . 355

*Subtle Biases Need Subtler Measures: Dual Metrics for Evaluating Representative and Affinity Bias in Large Language Models*
Abhishek Kumar, Sarfaroz Yunusov and Ali Emami . . . . . 375

*Framing in the Presence of Supporting Data: A Case Study in U.S. Economic News*
Alexandria Leto, Elliot E. Pickens, Coen D. Needell, David Rothschild and Maria Leonor Pacheco 393

*Mementos: A Comprehensive Benchmark for Multimodal Large Language Model Reasoning over Image Sequences*
Xiyao Wang, Yuhang Zhou, Xiaoyu Liu, Hongjin Lu, Yuancheng Xu, Feihong He, Jaehong Yoon, Taixi Lu, Fuxiao Liu, Gedas Bertasius, Mohit Bansal, Huaxiu Yao and Furong Huang . . . . . 416

*TTM-RE: Memory-Augmented Document-Level Relation Extraction*
Chufan Gao, Xuan Wang and Jimeng Sun . . . . . 443

*Answer is All You Need: Instruction-following Text Embedding via Answering the Question*
Letian Peng, Yuwei Zhang, Zilong Wang, Jayanth Srinivasa, Gaowen Liu, Zihan Wang and Jingbo Shang . . . . . 459

*Explore Spurious Correlations at the Concept Level in Language Models for Text Classification*
Yuhang Zhou, Paiheng Xu, Xiaoyu Liu, Bang An, Wei Ai and Furong Huang . . . . . 478

*Every Answer Matters: Evaluating Commonsense with Probabilistic Measures*
Qi Cheng, Michael Boratko, Pranay Kumar Yelugam, Tim O'Gorman, Nalini Singh, Andrew McCallum and Xiang Lorraine Li . . . . . 493

*GradSafe: Detecting Jailbreak Prompts for LLMs via Safety-Critical Gradient Analysis*
Yueqi Xie, Minghong Fang, Renjie Pi and Neil Zhenqiang Gong . . . . . 507

*Pouring Your Heart Out: Investigating the Role of Figurative Language in Online Expressions of Empathy*
Gyeongeun Lee, Christina Wong, Meghan Guo and Natalie Parde . . . . . 519

*An Information-Theoretic Approach to Analyze NLP Classification Tasks*
Luran Wang, Mark Gales and Vatsal Raina . . . . . 530

*Can Your Model Tell a Negation from an Implicature? Unravelling Challenges With Intent Encoders*
Yuwei Zhang, Siffi Singh, Sailik Sengupta, Igor Shalyminov, Hang Su, Hwanjun Song and Saab Mansour . . . . . 552

*Wav2Gloss: Generating Interlinear Glossed Text from Speech*
Taiqi He, Kwanghee Choi, Lindia Tjuatja, Nathaniel Romney Robinson, Jiatong Shi, Shinji Watanabe, Graham Neubig, David R Mortensen and Lori Levin . . . . . 568

*Leveraging Codebook Knowledge with NLI and ChatGPT for Zero-Shot Political Relation Classification*
Yibo Hu, Erick Skorupa Parolin, Latifur Khan, Patrick Brandt, Javier Osorio and Vito D'Orazio 583

*SPOR: A Comprehensive and Practical Evaluation Method for Compositional Generalization in Data-to-Text Generation*
Ziyao Xu and Houfeng Wang . . . . . 604

*OPEx: A Component-Wise Analysis of LLM-Centric Agents in Embodied Instruction Following*
Haochen Shi, Zhiyuan Sun, Xingdi Yuan, Marc-Alexandre Côté and Bang Liu . . . . . 622

*Multimodal Instruction Tuning with Conditional Mixture of LoRA*
Ying Shen, Zhiyang Xu, Qifan Wang, Yu Cheng, Wenpeng Yin and Lifu Huang . . . . . 637

*DocLens: Multi-aspect Fine-grained Medical Text Evaluation*
Yiqing Xie, Sheng Zhang, Hao Cheng, Pengfei Liu, Zelalem Gero, Cliff Wong, Tristan Naumann, Hoifung Poon and Carolyn Rose . . . . . 649

*FOFO: A Benchmark to Evaluate LLMs' Format-Following Capability*
Congying Xia, Chen Xing, Jiangshu Du, Xinyi Yang, Yihao Feng, Ran Xu, Wenpeng Yin and Caiming Xiong . . . . . 680

*Hyper-CL: Conditioning Sentence Representations with Hypernetworks*
Young Hyun Yoo, Jii Cha, Changhyeon Kim and Taeuk Kim . . . . . 700

*Analysis of Multi-Source Language Training in Cross-Lingual Transfer*
Seonghoon Lim, Taejun Yun, Jinhyeon Kim, Jihun Choi and Taeuk Kim . . . . . 712

*ABEX: Data Augmentation for Low-Resource NLU via Expanding Abstract Descriptions*
Sreyan Ghosh, Utkarsh Tyagi, Sonal Kumar, Chandra Kiran Reddy Evuru, Ramaneswaran S, S Sakshi and Dinesh Manocha . . . . . 726

*The Belebele Benchmark: a Parallel Reading Comprehension Dataset in 122 Language Variants*
Lucas Bandarkar, Davis Liang, Benjamin Muller, Mikel Artetxe, Satya Narayan Shukla, Donald Husa, Naman Goyal, Abhinandan Krishnan, Luke Zettlemoyer and Madian Khabsa . . . . . 749

*Learn from Failure: Fine-tuning LLMs with Trial-and-Error Data for Intuitionistic Propositional Logic Proving*
Chenyang An, Zhibo Chen, Qihao Ye, Emily First, Letian Peng, Jiayun Zhang, Zihan Wang, Sorin Lerner and Jingbo Shang . . . . . 776

*Interactive Text-to-Image Retrieval with Large Language Models: A Plug-and-Play Approach*
Saehyung Lee, Sangwon Yu, Junsung Park, Jihun Yi and Sungroh Yoon . . . . . 791

*IMBUE: Improving Interpersonal Effectiveness through Simulation and Just-in-time Feedback with Human-Language Model Interaction*
Inna Wanyin Lin, Ashish Sharma, Christopher Michael Rytting, Adam S Miner, Jina Suh and Tim Althoff . . . . . 810

*Token-wise Influential Training Data Retrieval for Large Language Models*
Huawei Lin, Jikai Long, Zhaozhuo Xu and Weijie Zhao . . . . . 841

*Tree-of-Counterfactual Prompting for Zero-Shot Stance Detection*
Maxwell Weinzierl and Sanda Harabagiu . . . . . 861

*VisualWebArena: Evaluating Multimodal Agents on Realistic Visual Web Tasks*
Jing Yu Koh, Robert Lo, Lawrence Keunho Jang, Vikram Duvvur, Ming Chong Lim, Po-Yu Huang, Graham Neubig, Shuyan Zhou, Russ Salakhutdinov and Daniel Fried . . . . . 881

*FineSurE: Fine-grained Summarization Evaluation using LLMs*
Hwanjun Song, Hang Su, Igor Shalyminov, Jason Cai and Saab Mansour . . . . . 906

*Tuning Large Multimodal Models for Videos using Reinforcement Learning from AI Feedback*
Daechul Ahn, Yura Choi, Youngjae Yu, Dongyeop Kang and Jonghyun Choi . . . . . 923

*Prompt Refinement with Image Pivot for Text-to-Image Generation*
Jingtao Zhan, Qingyao Ai, Yiqun Liu, Yingwei Pan, Ting Yao, Jiaxin Mao, Shaoping Ma and Tao Mei . . . . . 941

*Striking Gold in Advertising: Standardization and Exploration of Ad Text Generation*
Masato Mita, Soichiro Murakami, Akihiko Kato and Peinan Zhang . . . . . 955

*AbsInstruct: Eliciting Abstraction Ability from LLMs through Explanation Tuning with Plausibility Estimation*
Zhaowei Wang, Wei Fan, Qing Zong, Hongming Zhang, Sehyun Choi, Tianqing Fang, Xin Liu, Yangqiu Song, Ginny Wong and Simon See . . . . . 973

*Reflect-RL: Two-Player Online RL Fine-Tuning for LMs*
Runlong Zhou, Simon Shaolei Du and Beibin Li . . . . . 995

*Can ChatGPT's Performance be Improved on Verb Metaphor Detection Tasks? Bootstrapping and Combining Tacit Knowledge*
Cheng Yang, Puli Chen and Qingbao Huang . . . . . 1016

*Self-Distillation Bridges Distribution Gap in Language Model Fine-Tuning*
Zhaorui Yang, Tianyu Pang, Haozhe Feng, Han Wang, Wei Chen, Minfeng Zhu and Qian Liu 1028

*An Information Bottleneck Perspective for Effective Noise Filtering on Retrieval-Augmented Generation*
Kun Zhu, Xiaocheng Feng, Xiyuan Du, Yuxuan Gu, Weijiang Yu, Haotian Wang, Qianglong Chen, Zheng Chu, Jingchang Chen and Bing Qin . . . . . 1044

*RORA: Robust Free-Text Rationale Evaluation*
Zhengping Jiang, Yining Lu, Hanjie Chen, Daniel Khashabi, Benjamin Van Durme and Anqi Liu 1070

*Tell Me More! Towards Implicit User Intention Understanding of Language Model Driven Agents*
Cheng Qian, Bingxiang He, Zhong Zhuang, Jia Deng, Yujia Qin, Xin Cong, Zhong Zhang, Jie Zhou, Yankai Lin, Zhiyuan Liu and Maosong Sun . . . . . 1088

*InstructProtein: Aligning Human and Protein Language via Knowledge Instruction*
Zeyuan Wang, Qiang Zhang, Keyan Ding, Ming Qin, Xiang Zhuang, Xiaotong Li and Huajun Chen . . . . . 1114

*ConSiDERS-The-Human Evaluation Framework: Rethinking Human Evaluation for Generative Large Language Models*
Aparna Elangovan, Ling Liu, Lei Xu, Sravan Babu Bodapati and Dan Roth . . . . . 1137

*Linguistically Conditioned Semantic Textual Similarity*
Jingxuan Tu, Keer Xu, Liulu Yue, Bingyang Ye, Kyeongmin Rim and James Pustejovsky . . 1161

*Navigate through Enigmatic Labyrinth A Survey of Chain of Thought Reasoning: Advances, Frontiers and Future*
Zheng Chu, Jingchang Chen, Qianglong Chen, Weijiang Yu, Tao He, Haotian Wang, Weihua Peng, Ming Liu, Bing Qin and Ting Liu . . . . . 1173

*TimeBench: A Comprehensive Evaluation of Temporal Reasoning Abilities in Large Language Models*
Zheng Chu, Jingchang Chen, Qianglong Chen, Weijiang Yu, Haotian Wang, Ming Liu and Bing Qin . . . . . 1204

*BeamAggR: Beam Aggregation Reasoning over Multi-source Knowledge for Multi-hop Question Answering*
Zheng Chu, Jingchang Chen, Qianglong Chen, Haotian Wang, Kun Zhu, Xiyuan Du, Weijiang Yu, Ming Liu and Bing Qin . . . . . 1229

*ANALOGYKB: Unlocking Analogical Reasoning of Language Models with A Million-scale Knowledge Base*
Siyu Yuan, Jiangjie Chen, Changzhi Sun, Jiaqing Liang, Yanghua Xiao and Deqing Yang . . 1249

*TaSL: Continual Dialog State Tracking via Task Skill Localization and Consolidation*
Yujie Feng, Xu Chu, Yongxin Xu, Guangyuan Shi, Bo Liu and Xiao-Ming Wu . . . . . 1266

*DeepSeekMoE: Towards Ultimate Expert Specialization in Mixture-of-Experts Language Models*
Damai Dai, Chengqi Deng, Chenggang Zhao, R.x. Xu, Huazuo Gao, Deli Chen, Jiashi Li, Wangding Zeng, Xingkai Yu, Y. Wu, Zhenda Xie, Y.k. Li, Panpan Huang, Fuli Luo, Chong Ruan, Zhifang Sui and Wenfeng Liang . . . . . 1280

*Grounding Language Model with Chunking-Free In-Context Retrieval*
Hongjin Qian, Zheng Liu, Kelong Mao, Yujia Zhou and Zhicheng Dou . . . . . 1298

*Advancing Abductive Reasoning in Knowledge Graphs through Complex Logical Hypothesis Generation*
Jiaxin Bai, Yicheng Wang, Tianshi Zheng, Yue Guo, Xin Liu and Yangqiu Song . . . . . 1312

*Active Prompting with Chain-of-Thought for Large Language Models*
Shizhe Diao, Pengcheng Wang, Yong Lin, Rui Pan, Xiang Liu and Tong Zhang . . . . . 1330

*EasyGen: Easing Multimodal Generation with BiDiffuser and LLMs*
Xiangyu Zhao, Bo Liu, Qijiong Liu, Guangyuan Shi and Xiao-Ming Wu . . . . . 1351

*Rewriting the Code: A Simple Method for Large Language Model Augmented Code Search*
Haochen Li, Xin Zhou and Zhiqi Shen . . . . . 1371

*A Multidimensional Framework for Evaluating Lexical Semantic Change with Social Science Applications*
Naomi Baes, Nick Haslam and Ekaterina Vylomova . . . . . 1390

*Mitigating Catastrophic Forgetting in Large Language Models with Self-Synthesized Rehearsal*
Jianheng Huang, Leyang Cui, Ante Wang, Chengyi Yang, Xinting Liao, Linfeng Song, Junfeng Yao and Jinsong Su . . . . . 1416

*Enhancing Large Language Models in Coding Through Multi-Perspective Self-Consistency*
Baizhou Huang, Shuai Lu, Xiaojun Wan and Nan Duan . . . . . 1429

*Citation-Enhanced Generation for LLM-based Chatbots*
Weitao Li, Junkai Li, Weizhi Ma and Yang Liu . . . . . 1451

*Transitive Consistency Constrained Learning for Entity-to-Entity Stance Detection*
Haoyang Wen, Eduard Hovy and Alexander G Hauptmann . . . . . 1467

*Feature-Adaptive and Data-Scalable In-Context Learning*
Jiahao Li, Quan Wang, Licheng Zhang, Guoqing Jin and Zhendong Mao . . . . . 1481

*Probing the Multi-turn Planning Capabilities of LLMs via 20 Question Games*
Yizhe Zhang, Jiarui Lu and Navdeep Jaitly . . . . . 1495

*WaterBench: Towards Holistic Evaluation of Watermarks for Large Language Models*
Shangqing Tu, Yuliang Sun, Yushi Bai, Jifan Yu, Lei Hou and Juanzi Li . . . . . 1517

*Dependency Transformer Grammars: Integrating Dependency Structures into Transformer Language Models*
Yida Zhao, Chao Lou and Kewei Tu . . . . . 1543

*A Non-autoregressive Generation Framework for End-to-End Simultaneous Speech-to-Any Translation*
Zhengrui Ma, Qingkai Fang, Shaolei Zhang, Shoutao Guo, Yang Feng and Min Zhang . . . . . 1557

*Probing Language Models for Pre-training Data Detection*
Zhenhua Liu, Tong Zhu, Chuanyuan Tan, Bing Liu, Haonan Lu and Wenliang Chen . . . . . 1576

*Analyzing Temporal Complex Events with Large Language Models? A Benchmark towards Temporal, Long Context Understanding*
Zhihan Zhang, Yixin Cao, Chenchen Ye, Yunshan Ma, Lizi Liao and Tat-Seng Chua . . . . . 1588

*IBSEN: Director-Actor Agent Collaboration for Controllable and Interactive Drama Script Generation*
Senyu Han, Lu Chen, Li-Min Lin, Zhengshan Xu and Kai Yu . . . . . 1607

*Language Model Adaption for Reinforcement Learning with Natural Language Action Space*
Jiangxing Wang, Jiachen Li, Xiao Han, Deheng Ye and Zongqing Lu . . . . . 1620

*Evaluating Intention Detection Capability of Large Language Models in Persuasive Dialogues*
Hiromasa Sakurai and Yusuke Miyao . . . . . 1635

*LongLLMLingua: Accelerating and Enhancing LLMs in Long Context Scenarios via Prompt Compression*
Huiqiang Jiang, Qianhui Wu, Xufang Luo, Dongsheng Li, Chin-Yew Lin, Yuqing Yang and Lili Qiu . . . . . 1658

*Persuading across Diverse Domains: a Dataset and Persuasion Large Language Model*
Chuhao Jin, Kening Ren, Lingzhen Kong, Xiting Wang, Ruihua Song and Huan Chen . . . . . 1678

*HealMe: Harnessing Cognitive Reframing in Large Language Models for Psychotherapy*
Mengxi Xiao, Qianqian Xie, Ziyan Kuang, Zhicheng Liu, Kailai Yang, Min Peng, Weiguang Han and Jimin Huang . . . . . 1707

*Multimodal Prompt Learning with Missing Modalities for Sentiment Analysis and Emotion Recognition*
Zirun Guo, Tao Jin and Zhou Zhao . . . . . 1726

*An Effective Pronunciation Assessment Approach Leveraging Hierarchical Transformers and Pre-training Strategies*
Bi-Cheng Yan, Jiun-Ting Li, Yi-Cheng Wang, Hsin Wei Wang, Tien-Hong Lo, Yung-Chang Hsu, Wei-Cheng Chao and Berlin Chen . . . . . 1737

*Detection-Correction Structure via General Language Model for Grammatical Error Correction*
Wei Li and Houfeng Wang . . . . . 1748

*Generative Pre-trained Speech Language Model with Efficient Hierarchical Transformer*
Yongxin Zhu, Dan Su, Liqiang He, Linli Xu and Dong Yu . . . . . 1764

*Selene: Pioneering Automated Proof in Software Verification*
Lichen Zhang, Shuai Lu and Nan Duan . . . . . 1776

*Dissecting Human and LLM Preferences*
Junlong Li, Fan Zhou, Shichao Sun, Yikai Zhang, Hai Zhao and Pengfei Liu . . . . . 1790

*UniCoder: Scaling Code Large Language Model via Universal Code*
Tao Sun, Linzheng Chai, Jian Yang, Yuwei Yin, Hongcheng Guo, Jiaheng Liu, Bing Wang, Liqun Yang and Zhoujun Li . . . . . 1812

*AoE: Angle-optimized Embeddings for Semantic Textual Similarity*
Xianming LI and Jing Li . . . . . 1825

*InCharacter: Evaluating Personality Fidelity in Role-Playing Agents through Psychological Interviews*
Xintao Wang, Yunze Xiao, Jen-tse Huang, Siyu Yuan, Rui Xu, Haoran Guo, Quan Tu, Yaying Fei, Ziang Leng, Wei Wang, Jiangjie Chen, Cheng Li and Yanghua Xiao . . . . . 1840

*Does DetectGPT Fully Utilize Perturbation? Bridging Selective Perturbation to Fine-tuned Contrastive Learning Detector would be Better*
Shengchao Liu, Xiaoming Liu, Yichen Wang, Zehua Cheng, Chengzhengxu Li, Zhaohan Zhang, Yu Lan and Chao Shen . . . . . 1874

*AFaCTA: Assisting the Annotation of Factual Claim Detection with Reliable LLM Annotators*
Jingwei Ni, Minjing Shi, Dominik Stammbach, Mrinmaya Sachan, Elliott Ash and Markus Leippold . . . . . 1890

*Towards Faithful and Robust LLM Specialists for Evidence-Based Question-Answering*
Tobias Schimanski, Jingwei Ni, Mathias Kraus, Elliott Ash and Markus Leippold . . . . . 1913

*LoRAMoE: Alleviating World Knowledge Forgetting in Large Language Models via MoE-Style Plugin*
Shihan Dou, Enyu Zhou, Yan Liu, Songyang Gao, Wei Shen, Limao Xiong, Yuhao Zhou, Xiao Wang, Zhiheng Xi, Xiaoran Fan, Shiliang Pu, Jiang Zhu, Rui Zheng, Tao Gui, Qi Zhang and Xuanjing Huang . . . . . 1932

*Self-Alignment for Factuality: Mitigating Hallucinations in LLMs via Self-Evaluation*
Xiaoying Zhang, Baolin Peng, Ye Tian, Jingyan Zhou, Lifeng Jin, Linfeng Song, Haitao Mi and Helen M. Meng . . . . . 1946

*M-RAG: Reinforcing Large Language Model Performance through Retrieval-Augmented Generation with Multiple Partitions*
Zheng Wang, Shu Xian Teo, Jieer Ouyang, Yongjun xu and Wei Shi . . . . . 1966

*AIR-Bench: Benchmarking Large Audio-Language Models via Generative Comprehension*
Qian Yang, Jin Xu, Wenrui Liu, Yunfei Chu, Ziyue Jiang, Xiaohuan Zhou, Yichong Leng, Yuanjun Lv, Zhou Zhao, Chang Zhou and Jingren Zhou . . . . . 1979

*Navigating the Metrics Maze: Reconciling Score Magnitudes and Accuracies*
Tom Kocmi, Vilém Zouhar, Christian Federmann and Matt Post . . . . . 1999

*ValueBench: Towards Comprehensively Evaluating Value Orientations and Understanding of Large Language Models*
Yuanyi Ren, Haoran Ye, Hanjun Fang, Xin Zhang and Guojie Song . . . . . 2015

*DM-BLI: Dynamic Multiple Subspaces Alignment for Unsupervised Bilingual Lexicon Induction*
Ling Hu and Yuemei Xu . . . . . 2041

*SparseFit: Few-shot Prompting with Sparse Fine-tuning for Jointly Generating Predictions and Natural Language Explanations*
Jesus Solano, Mardhiyah Sanni, Oana-Maria Camburu and Pasquale Minervini . . . . . 2053

*Handling Ambiguity in Emotion: From Out-of-Domain Detection to Distribution Estimation*
Wen Wu, Bo Li, Chao Zhang, Chung-Cheng Chiu, Qiujia Li, Junwen Bai, Tara N Sainath and Phil Woodland . . . . . 2078

*REANO: Optimising Retrieval-Augmented Reader Models through Knowledge Graph Generation*
Jinyuan Fang, Zaiqiao Meng and Craig MacDonald . . . . . 2094

*Learning Disentangled Semantic Spaces of Explanations via Invertible Neural Networks*
Yingji Zhang, Danilo Carvalho and Andre Freitas . . . . . 2113

*MoPS: Modular Story Premise Synthesis for Open-Ended Automatic Story Generation*
Yan Ma, Yu Qiao and Pengfei Liu . . . . . 2135

*Open-Set Semi-Supervised Text Classification via Adversarial Disagreement Maximization*
Junfan Chen, Richong Zhang, Junchi Chen and Chunming Hu . . . . . 2170

*ToolSword: Unveiling Safety Issues of Large Language Models in Tool Learning Across Three Stages*
Junjie Ye, Sixian Li, Guanyu Li, Caishuang Huang, Songyang Gao, Yilong Wu, Qi Zhang, Tao Gui and Xuanjing Huang . . . . . 2181

*A synthetic data approach for domain generalization of NLI models*
Mohammad Javad Hosseini, Andrey Petrov, Alex Fabrikant and Annie Louis . . . . . 2212

*Enhancing Contrastive Learning with Noise-Guided Attack: Towards Continual Relation Extraction in the Wild*
Ting Wu, Jingyi Liu, Rui Zheng, Tao Gui, Qi Zhang and Xuanjing Huang . . . . . 2227

*LRQuant: Learnable and Robust Post-Training Quantization for Large Language Models*
Jiaqi Zhao, Miao Zhang, Chao Zeng, Ming Wang, Xuebo Liu and Liqiang Nie . . . . . 2240

*VariErr NLI: Separating Annotation Error from Human Label Variation*
Leon Weber-Genzel, Siyao Peng, Marie-Catherine De Marneffe and Barbara Plank . . . . . 2256

*Benchmarking Knowledge Boundary for Large Language Models: A Different Perspective on Model Evaluation*
Xunjian Yin, Xu Zhang, Jie Ruan and Xiaojun Wan . . . . . 2270

*ListT5: Listwise Reranking with Fusion-in-Decoder Improves Zero-shot Retrieval*
Soyoung Yoon, Eunbi Choi, Jiyeon Kim, Hyeongu Yun, Yireun Kim and Seung-won Hwang 2287

*Exploring the Potential of Large Language Models in Computational Argumentation*
Guizhen Chen, Liying Cheng, Anh Tuan Luu and Lidong Bing . . . . . 2309

*TaxoLLaMA: WordNet-based Model for Solving Multiple Lexical Semantic Tasks*
Viktor Moskvoretskii, Ekaterina Neminova, Alina Lobanova, Alexander Panchenko and Irina Nikishina . . . . . 2331

*CANDLE: Iterative Conceptualization and Instantiation Distillation from Large Language Models for Commonsense Reasoning*
Weiqi Wang, Tianqing Fang, Chunyang Li, Haochen Shi, Wenxuan Ding, Baixuan Xu, Zhaowei Wang, Jiaxin Bai, Xin Liu, Cheng Jiayang, Chunkit Chan and Yangqiu Song . . . . . 2351

*MEFT: Memory-Efficient Fine-Tuning through Sparse Adapter*
Jitai Hao, Weiwei Sun, Xin Xin, Qi Meng, Zhumin Chen, Pengjie Ren and Zhaochun Ren . . 2375

*Surgical Feature-Space Decomposition of LLMs: Why, When and How?*
Arnav Chavan, Nahush Lele and Deepak Gupta . . . . . 2389

*Reasoning in Flux: Enhancing Large Language Models Reasoning through Uncertainty-aware Adaptive Guidance*
Zhangyue Yin, Qiushi Sun, Qipeng Guo, Zhiyuan Zeng, Xiaonan Li, Junqi Dai, Qinyuan Cheng, Xuanjing Huang and Xipeng Qiu . . . . . 2401

*Modality-Aware Integration with Large Language Models for Knowledge-Based Visual Question Answering*
Junnan Dong, Qinggang Zhang, Huachi Zhou, Daochen Zha, Pai Zheng and Xiao Huang . . 2417

*Unlocking Data-free Low-bit Quantization with Matrix Decomposition for KV Cache Compression*
Peiyu Liu, Ze-Feng Gao, Xin Zhao, Yipeng Ma, Tao Wang and Ji-Rong Wen . . . . . 2430

*VerifiNER: Verification-augmented NER via Knowledge-grounded Reasoning with Large Language Models*
Seoyeon Kim, Kwangwook Seo, Hyungjoo Chae, Jinyoung Yeo and Dongha Lee . . . . . 2441

*Making Long-Context Language Models Better Multi-Hop Reasoners*
Yanyang Li, Shuo Liang, Michael Lyu and Liwei Wang . . . . . 2462

*TransliCo: A Contrastive Learning Framework to Address the Script Barrier in Multilingual Pretrained Language Models*
Yihong Liu, Chunlan Ma, Haotian Ye and Hinrich Schuetze . . . . . 2476

*Extreme Miscalibration and the Illusion of Adversarial Robustness*
Vyas Raina, Samson Tan, Volkan Cevher, Aditya Rawal, Sheng Zha and George Karypis . . 2500

*HyCoRec: Hypergraph-Enhanced Multi-Preference Learning for Alleviating Matthew Effect in Conversational Recommendation*
Yongsen Zheng, Ruilin Xu, Ziliang Chen, Guohua Wang, Mingjie Qian, Jinghui Qin and Liang Lin . . . . . 2526

*Co-training for Low Resource Scientific Natural Language Inference*
Mobashir Sadat and Cornelia Caragea . . . . . 2538

*RLHFPoison: Reward Poisoning Attack for Reinforcement Learning with Human Feedback in Large Language Models*
Jiongxiao Wang, Junlin Wu, Muhao Chen, Yevgeniy Vorobeychik and Chaowei Xiao . . . . . 2551

*Time is Encoded in the Weights of Finetuned Language Models*
Kai Nylund, Suchin Gururangan and Noah A. Smith . . . . . 2571

*Long-Context Language Modeling with Parallel Context Encoding*
Howard Yen, Tianyu Gao and Danqi Chen . . . . . 2588

*SirLLM: Streaming Infinite Retentive LLM*
Yao Yao, Zuchao Li and Hai Zhao . . . . . 2611

*IMO: Greedy Layer-Wise Sparse Representation Learning for Out-of-Distribution Text Classification with Pre-trained Models*
Tao Feng, Lizhen Qu, Zhuang Li, Haolan Zhan, Yuncheng Hua and Reza Haf . . . . . 2625

*Generative Pretrained Structured Transformers: Unsupervised Syntactic Language Models at Scale*
Xiang Hu, Pengyu Ji, Qingyang Zhu, Wei Wu and Kewei Tu . . . . . 2640

*MELA: Multilingual Evaluation of Linguistic Acceptability*
Ziyin Zhang, Yikang Liu, Weifang Huang, Junyu Mao, Rui Wang and Hai Hu . . . . . 2658

*CopyNE: Better Contextual ASR by Copying Named Entities*
Shilin Zhou, Zhenghua Li, Yu Hong, Min Zhang, Zhefeng Wang and Baoxing Huai . . . . . 2675

*Is Table Retrieval a Solved Problem? Exploring Join-Aware Multi-Table Retrieval*
Peter Baile Chen, Yi Zhang and Dan Roth . . . . . 2687

*Generalizing Conversational Dense Retrieval via LLM-Cognition Data Augmentation*
Haonan Chen, Zhicheng Dou, Kelong Mao, Jiongnan Liu and Ziliang Zhao . . . . . 2700

*ItD: Large Language Models Can Teach Themselves Induction through Deduction*
Wangtao Sun, Haotian Xu, Xuanqing Yu, Pei Chen, Shizhu He, Jun Zhao and Kang Liu . . . . 2719

*MathGenie: Generating Synthetic Data with Question Back-translation for Enhancing Mathematical Reasoning of LLMs*
Zimu Lu, Aojun Zhou, Houxing Ren, Ke Wang, Weikang Shi, Junting Pan, Mingjie Zhan and Hongsheng Li . . . . . 2732

*Rethinking Task-Oriented Dialogue Systems: From Complex Modularity to Zero-Shot Autonomous Agent*
Heng-Da Xu, Xian-Ling Mao, Puhai Yang, Fanshu Sun and Heyan Huang . . . . . 2748

*On Context Utilization in Summarization with Large Language Models*
Mathieu Ravaut, Aixin Sun, Nancy F. Chen and Shafiq Joty . . . . . 2764

*INTERS: Unlocking the Power of Large Language Models in Search with Instruction Tuning*
Yutao Zhu, Peitian Zhang, Chenghao Zhang, Yifei Chen, Binyu Xie, Zheng Liu, Ji-Rong Wen and Zhicheng Dou . . . . . 2782

*Enhancing In-Context Learning via Implicit Demonstration Augmentation*
Xiaoling Zhou, Wei Ye, Yidong Wang, Chaoya Jiang, Zhemg Lee, Rui Xie and Shikun Zhang 2810

*PRoLoRA: Partial Rotation Empowers More Parameter-Efficient LoRA*
Sheng Wang, Boyang Xue, Jiacheng Ye, Jiyue Jiang, Liheng Chen, Lingpeng Kong and Chuan Wu . . . . . 2829

*Improving Event Definition Following For Zero-Shot Event Detection*
Zefan Cai, Po-Nien Kung, Ashima Suvarna, Mingyu Derek Ma, Hritik Bansal, Baobao Chang, P. Jeffrey Brantingham, Wei Wang and Nanyun Peng . . . . . 2842

*Through the MUD: A Multi-Defendant Charge Prediction Benchmark with Linked Crime Elements*
Xiao Wei, Qi Xu, Hang Yu, Qian Liu and Erik Cambria . . . . . 2864

*Interpreting Conversational Dense Retrieval by Rewriting-Enhanced Inversion of Session Embedding*
Yiruo Cheng, Kelong Mao and Zhicheng Dou . . . . . 2879

*Stumbling Blocks: Stress Testing the Robustness of Machine-Generated Text Detectors Under Attacks*
Yichen Wang, Shangbin Feng, Abe Bohan Hou, Xiao Pu, Chao Shen, Xiaoming Liu, Yulia Tsvetkov and Tianxing He . . . . . 2894

*Training Language Models to Generate Text with Citations via Fine-grained Rewards*
Chengyu Huang, Zeqiu Wu, Yushi Hu and Wenya Wang . . . . . 2926

*Hypergraph based Understanding for Document Semantic Entity Recognition*
Qiwei Li, Zuchao Li, Ping Wang, Haojun Ai and Hai Zhao . . . . . 2950

*GSM-Plus: A Comprehensive Benchmark for Evaluating the Robustness of LLMs as Mathematical Problem Solvers*
Qintong Li, Leyang Cui, Xueliang Zhao, Lingpeng Kong and Wei Bi . . . . . 2961

*Synergetic Event Understanding: A Collaborative Approach to Cross-Document Event Coreference Resolution with Large Language Models*
Qingkai Min, Qipeng Guo, Xiangkun Hu, Songfang Huang, Zheng Zhang and Yue Zhang . . 2985

*AutoAct: Automatic Agent Learning from Scratch for QA via Self-Planning*
Shuofei Qiao, Ningyu Zhang, Runnan Fang, Yujie Luo, Wangchunshu Zhou, Yuchen Eleanor Jiang, Chengfei lv and Huajun Chen . . . . . 3003

*ChronosLex: Time-aware Incremental Training for Temporal Generalization of Legal Classification Tasks*
Santosh T.y.s.s, Tuan-Quang Vuong and Matthias Grabmair . . . . . 3022

*Virtual Compiler Is All You Need For Assembly Code Search*
Zeyu Gao, Hao Wang, Yuanda Wang and Chao Zhang . . . . . 3040

*MELoRA: Mini-Ensemble Low-Rank Adapters for Parameter-Efficient Fine-Tuning*
Pengjie Ren, Chengshun Shi, Shiguang Wu, Mengqi Zhang, Zhaochun Ren, Maarten de Rijke, Zhumin Chen and Jiahuan Pei . . . . . 3052

*Can LLMs Learn from Previous Mistakes? Investigating LLMs' Errors to Boost for Reasoning*
Yongqi Tong, Dawei Li, Sizhe Wang, Yujia Wang, Fei Teng and Jingbo Shang . . . . . 3065

*An Iterative Associative Memory Model for Empathetic Response Generation*
Zhou Yang, Zhaochun Ren, Wang Yufeng, Haizhou Sun, Chao Chen, Xiaofei Zhu and Xiangwen Liao . . . . . 3081

*Detoxifying Large Language Models via Knowledge Editing*
Mengru Wang, Ningyu Zhang, Ziwen Xu, Zekun Xi, Shumin Deng, Yunzhi Yao, Qishen Zhang, Linyi Yang, Jindong Wang and Huajun Chen . . . . . 3093

*LongBench: A Bilingual, Multitask Benchmark for Long Context Understanding*
Yushi Bai, Xin Lv, Jiajie Zhang, Hongchang Lyu, Jiankai Tang, Zhidian Huang, Zhengxiao Du, Xiao Liu, Aohan Zeng, Lei Hou, Yuxiao Dong, Jie Tang and Juanzi Li . . . . . 3119

*Dr.Academy: A Benchmark for Evaluating Questioning Capability in Education for Large Language Models*
Yuyan Chen, Songzhou Yan, Panjun Liu and Yanghua Xiao . . . . . 3138

*UniBridge: A Unified Approach to Cross-Lingual Transfer Learning for Low-Resource Languages*
Trinh Pham, Khoi M. Le and Anh Tuan Luu . . . . . 3168

*VISTA: Visualized Text Embedding For Universal Multi-Modal Retrieval*
Junjie Zhou, Zheng Liu, Shitao Xiao, Bo Zhao and Yongping Xiong . . . . . 3185

*Black-Box Prompt Optimization: Aligning Large Language Models without Model Training*
Jiale Cheng, Xiao Liu, Kehan Zheng, Pei Ke, Hongning Wang, Yuxiao Dong, Jie Tang and Minlie Huang . . . . . 3201

*Open Ko-LLM Leaderboard: Evaluating Large Language Models in Korean with Ko-H5 Benchmark*
Chanjun Park, Hyeonwoo Kim, Dahyun Kim, SeongHwan Cho, Sanghoon Kim, Sukyung Lee, Yungi Kim and Hwalsuk Lee . . . . . 3220

*Unified Hallucination Detection for Multimodal Large Language Models*
Xiang Chen, Chenxi Wang, Yida Xue, Ningyu Zhang, Xiaoyan Yang, Qiang Li, Yue Shen, Lei Liang, Jinjie GU and Huajun Chen . . . . . 3235

*Empowering Character-level Text Infilling by Eliminating Sub-Tokens*
Houxing Ren, Mingjie Zhan, Zhongyuan Wu and Hongsheng Li . . . . . 3253

*Landmark Embedding: A Chunking-Free Embedding Method For Retrieval Augmented Long-Context Large Language Models*
Kun Luo, Zheng Liu, Shitao Xiao, Tong Zhou, Yubo Chen, Jun Zhao and Kang Liu . . . . . 3268

*GrowOVER: How Can LLMs Adapt to Growing Real-World Knowledge?*
Dayoon Ko, Jinyoung Kim, Hahyeon Choi and Gunhee Kim . . . . . 3282

*Attribute First, then Generate: Locally-attributable Grounded Text Generation*
Aviv Slobodkin, Eran Hirsch, Arie Cattan, Tal Schuster and Ido Dagan . . . . . 3309

*T2S-GPT: Dynamic Vector Quantization for Autoregressive Sign Language Production from Text*
Aoxiong Yin, Haoyuan Li, Kai Shen, Siliang Tang and Yueting Zhuang . . . . . 3345

*OceanGPT: A Large Language Model for Ocean Science Tasks*
Zhen Bi, Ningyu Zhang, Yida Xue, Yixin Ou, Daxiong Ji, Guozhou Zheng and Huajun Chen 3357

*Beyond Memorization: The Challenge of Random Memory Access in Language Models*
Tongyao Zhu, Qian Liu, Liang Pang, Zhengbao Jiang, Min-Yen Kan and Min Lin . . . . . 3373

*BIPED: Pedagogically Informed Tutoring System for ESL Education*
Soonwoo Kwon, Sojung Kim, Minju Park, Seunghyun Lee and Kyuseok Kim . . . . . 3389

*Timeline-based Sentence Decomposition with In Context Learning for Temporal Fact Extraction*
Jianhao Chen, Haoyuan Ouyang, Junyang Ren, Wentao Ding, Wei Hu and Yuzhong Qu . . . . 3415

*Collaboration or Corporate Capture? Quantifying NLP's Reliance on Industry Artifacts and Contributions*
Will Aitken, Mohamed Abdalla, Karen Rudie and Catherine Stinson . . . . . 3433

*Prompt Expansion for Adaptive Text-to-Image Generation*
Siddhartha Datta, Alexander Ku, Deepak Ramachandran and Peter Anderson . . . . . 3449

*Progressively Modality Freezing for Multi-Modal Entity Alignment*
Yani Huang, Xuefeng Zhang, Richong Zhang, Junfan Chen and Jaein Kim . . . . . 3477

*Llama2Vec: Unsupervised Adaptation of Large Language Models for Dense Retrieval*
Chaofan Li, Zheng Liu, Shitao Xiao, Yingxia Shao and Defu Lian . . . . . 3490

*Democratizing LLMs for Low-Resource Languages by Leveraging their English Dominant Abilities with Linguistically-Diverse Prompts*
Xuan-Phi Nguyen, Mahani Aljunied, Shafiq Joty and Lidong Bing . . . . . 3501

*Metaphor Understanding Challenge Dataset for LLMs*
Xiaoyu Tong, Rochelle Choenni, Martha Lewis and Ekaterina Shutova . . . . . 3517

*A Multi-Task Embedder For Retrieval Augmented LLMs*
Peitian Zhang, Zheng Liu, Shitao Xiao, Zhicheng Dou and Jian-Yun Nie . . . . . 3537

*Language Models Don't Learn the Physical Manifestation of Language*
Bruce W. Lee and Jaehyuk Lim . . . . . 3554

*What Does the Bot Say? Opportunities and Risks of Large Language Models in Social Media Bot Detection*
Shangbin Feng, Herun Wan, Ningnan Wang, Zhaoxuan Tan, Minnan Luo and Yulia Tsvetkov 3580

*Self-Contrast: Better Reflection Through Inconsistent Solving Perspectives*
Wenqi Zhang, Yongliang Shen, Linjuan Wu, Qiuying Peng, Jun Wang, Yueting Zhuang and Weiming Lu . . . . . 3602

*Relying on the Unreliable: The Impact of Language Models' Reluctance to Express Uncertainty*
Kaitlyn Zhou, Jena D. Hwang, Xiang Ren and Maarten Sap . . . . . 3623

*Unity in Diversity: Collaborative Pre-training Across Multimodal Medical Sources*
Xiaochen Wang, Junyu Luo, Jiaqi Wang, Yuan Zhong, Xiaokun Zhang, Yaqing Wang, Parminder Bhatia, Cao Xiao and Fenglong Ma . . . . . 3644

*When Good and Reproducible Results are a Giant with Feet of Clay: The Importance of Software Quality in NLP*
Sara Papi, Marco Gaido, Andrea Pilzer and Matteo Negri . . . . . 3657

*SBAAM! Eliminating Transcript Dependency in Automatic Subtitling*
Marco Gaido, Sara Papi, Matteo Negri, Mauro Cettolo and Luisa Bentivogli . . . . . 3673

*StreamAtt: Direct Streaming Speech-to-Text Translation with Attention-based Audio History Selection*
Sara Papi, Marco Gaido, Matteo Negri and Luisa Bentivogli . . . . . 3692

*ARL2: Aligning Retrievers with Black-box Large Language Models via Self-guided Adaptive Relevance Labeling*
LingXi Zhang, Yue Yu, Kuan Wang and Chao Zhang . . . . . 3708

*Crayon: Customized On-Device LLM via Instant Adapter Blending and Edge-Server Hybrid Inference*
Jihwan Bang, Juntae Lee, Kyuhong Shim, Seunghan Yang and Simyung Chang . . . . . 3720

*FLEUR: An Explainable Reference-Free Evaluation Metric for Image Captioning Using a Large Multimodal Model*
Yebin Lee, Imseong Park and Myungjoo Kang . . . . . 3732

*MentalManip: A Dataset For Fine-grained Analysis of Mental Manipulation in Conversations*
Yuxin Wang, Ivory Yang, Saeed Hassanpour and Soroush Vosoughi . . . . . 3747

*MPCoder: Multi-user Personalized Code Generator with Explicit and Implicit Style Representation Learning*
Zhenlong Dai, Chang Yao, WenKang Han, Yuanying Yuanying, Zhipeng Gao and Jingyuan Chen 3765

*DataDreamer: A Tool for Synthetic Data Generation and Reproducible LLM Workflows*
Ajay Patel, Colin Raffel and Chris Callison-Burch . . . . . 3781

*Understanding and Addressing the Under-Translation Problem from the Perspective of Decoding Objective*
Chenze Shao, Fandong Meng, Jiali Zeng and Jie Zhou . . . . . 3800

*Identifying while Learning for Document Event Causality Identification*
Cheng Liu, Wei Xiang and Bang Wang . . . . . 3815

*OlympiadBench: A Challenging Benchmark for Promoting AGI with Olympiad-Level Bilingual Multimodal Scientific Problems*
Chaoqun He, Renjie Luo, Yuzhuo Bai, Shengding Hu, Zhen Leng Thai, Junhao Shen, Jinyi Hu, Xu Han, Yujie Huang, Yuxiang Zhang, Jie Liu, Lei Qi, Zhiyuan Liu and Maosong Sun . . . . . 3828

*Insert or Attach: Taxonomy Completion via Box Embedding*
Wei Xue, Yongliang Shen, Wenqi Ren, Jietian Guo, Shiliang Pu and Weiming Lu . . . . . 3851

*Semiparametric Token-Sequence Co-Supervision*
Hyunji Lee, Doyoung Kim, Jihoon Jun, Se June Joo, Joel Jang, Kyoung-Woon On and Minjoon Seo . . . . . 3864

*Instruction Fusion: Advancing Prompt Evolution through Hybridization*
Weidong Guo, Jiuding Yang, Kaitong Yang, Xiangyang Li, Zhuwei Rao, Yu Xu and Di Niu 3883

*TimeArena: Shaping Efficient Multitasking Language Agents in a Time-Aware Simulation*
Yikai Zhang, Siyu Yuan, Caiyu Hu, Kyle Richardson, Yanghua Xiao and Jiangjie Chen . . . . 3894

*Exploring Memorization in Fine-tuned Language Models*
Shenglai Zeng, Yaxin Li, Jie Ren, Yiding Liu, Han Xu, Pengfei He, Yue Xing, Shuaiqiang Wang, Jiliang Tang and Dawei Yin . . . . . 3917

*Towards Real-world Scenario: Imbalanced New Intent Discovery*
Shun Zhang, Yan Chaoran, Jian Yang, Jiaheng Liu, Ying Mo, Jiaqi Bai, Tongliang Li and Zhoujun Li . . . . . 3949

*M4GT-Bench: Evaluation Benchmark for Black-Box Machine-Generated Text Detection*
Yuxia Wang, Jonibek Mansurov, Petar Ivanov, Jinyan Su, Artem Shelmanov, Akim Tsvigun, Osama Mohammed Afzal, Tarek Mahmoud, Giovanni Puccetti, Thomas Arnold, Alham Fikri Aji, Nizar Habash, Iryna Gurevych and Preslav Nakov . . . . . 3964

*Instruct Once, Chat Consistently in Multiple Rounds: An Efficient Tuning Framework for Dialogue*
Jian Wang, Chak Tou Leong, Jiashuo Wang, Dongding Lin, Wenjie Li and Xiaoyong Wei . . 3993

*SoftDedup: an Efficient Data Reweighting Method for Speeding Up Language Model Pre-training*
Nan He, Weichen Xiong, Hanwen Liu, Yi Liao, Lei Ding, Kai Zhang, Guohua Tang, Xiao Han and Yang Wei . . . . . 4011

*Rule or Story, Which is a Better Commonsense Expression for Talking with Large Language Models?*
Ning Bian, Xianpei Han, Hongyu Lin, Yaojie Lu, Ben He and Le Sun . . . . . 4023

*Learning Global Controller in Latent Space for Parameter-Efficient Fine-Tuning*
Zeqi Tan, Yongliang Shen, Xiaoxia Cheng, Chang Zong, Wenqi Zhang, Jian Shao, Weiming Lu and Yueting Zhuang . . . . . 4044

*CaMML: Context-Aware Multimodal Learner for Large Models*
Yixin Chen, Shuai Zhang, Boran Han, Tong He and Bo Li . . . . . 4056

*MAVEN-ARG: Completing the Puzzle of All-in-One Event Understanding Dataset with Event Argument Annotation*
Xiaozhi Wang, Hao Peng, Yong Guan, Kaisheng Zeng, Jianhui Chen, Lei Hou, Xu Han, Yankai Lin, Zhiyuan Liu, Ruobing Xie, Jie Zhou and Juanzi Li . . . . . 4072

*NPHardEval: Dynamic Benchmark on Reasoning Ability of Large Language Models via Complexity Classes*
Lizhou Fan, Wenyue Hua, Lingyao Li, Haoyang Ling and Yongfeng Zhang . . . . . 4092

*Can Watermarks Survive Translation? On the Cross-lingual Consistency of Text Watermark for Large Language Models*
Zhiwei He, Binglin Zhou, Hongkun Hao, Aiwei Liu, Xing Wang, Zhaopeng Tu, Zhuosheng Zhang and Rui Wang . . . . . 4115

*Multi-Level Feedback Generation with Large Language Models for Empowering Novice Peer Counselors*
Alicja Chaszczewicz, Raj Sanjay Shah, Ryan Louie, Bruce A Arnow, Robert Kraut and Diyi Yang 4130

*In-context Mixing (ICM): Code-mixed Prompts for Multilingual LLMs*
Bhavani Shankar, Preethi Jyothi and Pushpak Bhattacharyya . . . . . 4162

*Respond in my Language: Mitigating Language Inconsistency in Response Generation based on Large Language Models*
Liang Zhang, Qin Jin, Haoyang Huang, Dongdong Zhang and Furu Wei . . . . . 4177

*Transferable Embedding Inversion Attack: Uncovering Privacy Risks in Text Embeddings without Model Queries*
Yu-Hsiang Huang, Yuche Tsai, Hsiang Hsiao, Hong-Yi Lin and Shou-De Lin . . . . . 4193

*Enhancing Reinforcement Learning with Label-Sensitive Reward for Natural Language Understanding*
Kuo Liao, Shuang Li, Meng Zhao, Liqun Liu, Mengge Xue, Zhenyu hu, Honglin Han and Chengguo Yin . . . . . 4206

*Intuitive or Dependent? Investigating LLMs' Behavior Style to Conflicting Prompts*
Jiahao Ying, Yixin Cao, Kai Xiong, Long Cui, Yidong He and Yongbin Liu . . . . . 4221

*CoCA: Fusing Position Embedding with Collinear Constrained Attention in Transformers for Long Context Window Extending*
Shiyi Zhu, Jing Ye, Wei Jiang, Siqiao Xue, Qi Zhang, Yifan Wu and Jianguo Li . . . . . 4247

*InfoLossQA: Characterizing and Recovering Information Loss in Text Simplification*
Jan Trienes, Sebastian Antony Joseph, Jörg Schlötterer, Christin Seifert, Kyle Lo, Wei Xu, Byron C Wallace and Junyi Jessy Li . . . . . 4263

*CoGenesis: A Framework Collaborating Large and Small Language Models for Secure Context-Aware Instruction Following*
Kaiyan Zhang, Jianyu Wang, Ermo Hua, Biqing Qi, Ning Ding and Bowen Zhou . . . . . 4295

*DAPR: A Benchmark on Document-Aware Passage Retrieval*
Kexin Wang, Nils Reimers and Iryna Gurevych . . . . . 4313

*Strengthened Symbol Binding Makes Large Language Models Reliable Multiple-Choice Selectors*
Mengge Xue, Zhenyu hu, Liqun Liu, Kuo Liao, Shuang Li, Honglin Han, Meng Zhao and Chengguo Yin . . . . . 4331

*SAC-KG: Exploiting Large Language Models as Skilled Automatic Constructors for Domain Knowledge Graph*
Hanzhu Chen, Xu Shen, Qitan Lv, Jie Wang, Xiaoqi Ni and Jieping Ye . . . . . 4345

*Uncertainty-Guided Modal Rebalance for Hateful Memes Detection*
Chuanpeng Yang, Yaxin Liu, Fuqing Zhu, Jizhong Han and Songlin Hu . . . . . 4361

*Missci: Reconstructing Fallacies in Misrepresented Science*
Max Glockner, Yufang Hou, Preslav Nakov and Iryna Gurevych . . . . . 4372

*Uncovering the Full Potential of Visual Grounding Methods in VQA*
Daniel Reich and Tanja Schultz . . . . . 4406

*Small Models, Big Insights: Leveraging Slim Proxy Models To Decide When and What to Retrieve for LLMs*
Jiejun Tan, Zhicheng Dou, Yutao Zhu, Peidong Guo, Kun Fang and Ji-Rong Wen . . . . . 4420

*Favi-Score: A Measure for Favoritism in Automated Preference Ratings for Generative AI Evaluation*
Pius Von Däniken, Jan Milan Deriu, Don Tuggener and Mark Cieliebak . . . . . 4437

*LLM-based Rewriting of Inappropriate Argumentation using Reinforcement Learning from Machine Feedback*
Timon Ziegenbein, Gabriella Skitalinskaya, Alireza Bayat Makou and Henning Wachsmuth 4455

*Graph Language Models*
Moritz Plenz and Anette Frank . . . . . 4477

*Analyzing Semantic Change through Lexical Replacements*
Francesco Periti, Pierluigi Cassotti, Haim Dubossarsky and Nina Tahmasebi . . . . . 4495

*Exploiting Intrinsic Multilateral Logical Rules for Weakly Supervised Natural Language Video Localization*
Zhe Xu, Kun Wei, Xu Yang and Cheng Deng . . . . . 4511

*Interpretability of Language Models via Task Spaces*
Lucas Weber, Jaap Jumelet, Elia Bruni and Dieuwke Hupkes . . . . . 4522

*Using Synchronic Definitions and Semantic Relations to Classify Semantic Change Types*
Pierluigi Cassotti, Stefano De Pascale and Nina Tahmasebi . . . . . 4539

*Factual Confidence of LLMs: on Reliability and Robustness of Current Estimators*
Matéo Mahaut, Laura Aina, Paula Czarnowska, Momchil Hardalov, Thomas Müller and Lluis Marquez . . . . . 4554

*StepCoder: Improving Code Generation with Reinforcement Learning from Compiler Feedback*
Shihan Dou, Yan Liu, Haoxiang Jia, Enyu Zhou, Limao Xiong, Junjie Shan, Caishuang Huang, Xiao Wang, Xiaoran Fan, Zhiheng Xi, Yuhao Zhou, Tao Ji, Rui Zheng, Qi Zhang, Tao Gui and Xuanjing Huang . . . . . 4571

*One-Shot Learning as Instruction Data Prospector for Large Language Models*
Yunshui Li, Binyuan Hui, Xiaobo Xia, Jiaxi Yang, Min Yang, Lei Zhang, Shuzheng Si, Ling-Hao Chen, Junhao Liu, Tongliang Liu, Fei Huang and Yongbin Li . . . . . 4586

*Navigating the OverKill in Large Language Models*
Chenyu Shi, Xiao Wang, Qiming Ge, Songyang Gao, Xianjun Yang, Tao Gui, Qi Zhang, Xuanjing Huang, Xun Zhao and Dahua Lin . . . . . 4602

*A Chain-of-Thought Is as Strong as Its Weakest Link: A Benchmark for Verifiers of Reasoning Chains*
Alon Jacovi, Yonatan Bitton, Bernd Bohnet, Jonathan Herzig, Or Honovich, Michael Tseng, Michael Collins, Roee Aharoni and Mor Geva . . . . . 4615

*Re3: A Holistic Framework and Dataset for Modeling Collaborative Document Revision*
Qian Ruan, Ilia Kuznetsov and Iryna Gurevych . . . . . 4635

*NextLevelBERT: Masked Language Modeling with Higher-Level Representations for Long Documents*
Tamara Czinczoll, Christoph Hönes, Maximilian Schall and Gerard De Melo . . . . . 4656

*FollowBench: A Multi-level Fine-grained Constraints Following Benchmark for Large Language Models*
Yuxin Jiang, Yufei Wang, Xingshan Zeng, Wanjun Zhong, Liangyou Li, Fei Mi, Lifeng Shang, Xin Jiang, Qun Liu and Wei Wang . . . . . 4667

*Learning to Edit: Aligning LLMs with Knowledge Editing*
Yuxin Jiang, Yufei Wang, Chuhan Wu, Wanjun Zhong, Xingshan Zeng, Jiahui Gao, Liangyou Li, Xin Jiang, Lifeng Shang, Ruiming Tang, Qun Liu and Wei Wang . . . . . 4689

*DolphCoder: Echo-Locating Code Large Language Models with Diverse and Multi-Objective Instruction Tuning*
Yejie Wang, Keqing He, Guanting Dong, Pei Wang, Weihao Zeng, Muxi Diao, Weiran Xu, Jingang Wang, Mengdi Zhang and Xunliang Cai . . . . . 4706

*When Only Time Will Tell: Interpreting How Transformers Process Local Ambiguities Through the Lens of Restart-Incrementality*
Brielen Madureira, Patrick Kahardipraja and David Schlangen . . . . . 4722

*SpaRC and SpaRP: Spatial Reasoning Characterization and Path Generation for Understanding Spatial Reasoning Capability of Large Language Models*
Md Imbesat Hassan Rizvi, Xiaodan Zhu and Iryna Gurevych . . . . . 4750

*Planning Like Human: A Dual-process Framework for Dialogue Planning*
Tao He, Lizi Liao, Yixin Cao, Yuanxing Liu, Ming Liu, Zerui Chen and Bing Qin . . . . . 4768

*Spectral Filters, Dark Signals, and Attention Sinks*
Nicola Cancedda . . . . . 4792

*DiffuCOMET: Contextual Commonsense Knowledge Diffusion*
Silin Gao, Mete Ismayilzada, Mengjie Zhao, Hiromi Wakaki, Yuki Mitsufuji and Antoine Bosselut 4809

*Systematic Task Exploration with LLMs: A Study in Citation Text Generation*
Furkan Şahinuç, Ilia Kuznetsov, Yufang Hou and Iryna Gurevych . . . . . 4832

*Limits of Theory of Mind Modelling in Dialogue-Based Collaborative Plan Acquisition*
Matteo Bortoletto, Constantin Ruhdorfer, Adnen Abdessaied, Lei Shi and Andreas Bulling . 4856

*Temporal Knowledge Question Answering via Abstract Reasoning Induction*
Ziyang Chen, Dongfang Li, Xiang Zhao, Baotian Hu and Min Zhang . . . . . 4872

*Who Wrote this Code? Watermarking for Code Generation*
Taehyun Lee, Seokhee Hong, Jaewoo Ahn, Ilgee Hong, Hwaran Lee, Sangdoo Yun, Jamin Shin and Gunhee Kim . . . . . 4890

*MapCoder: Multi-Agent Code Generation for Competitive Problem Solving*
Md. Ashraful Islam, Mohammed Eunus Ali and Md Rizwan Parvez . . . . . 4912

*RelayAttention for Efficient Large Language Model Serving with Long System Prompts*
Lei Zhu, Xinjiang Wang, Wayne Zhang and Rynson W. H. Lau . . . . . 4945

*Boosting Language Models Reasoning with Chain-of-Knowledge Prompting*
Jianing Wang, Qiushi Sun, Xiang Li and Ming Gao . . . . . 4958

*Open Grounded Planning: Challenges and Benchmark Construction*
Shiguang Guo, Ziliang Deng, Hongyu Lin, Yaojie Lu, Xianpei Han and Le Sun . . . . . 4982

*LLM Knows Body Language, Too: Translating Speech Voices into Human Gestures*
Chenghao Xu, Guangtao Lyu, Jiexi Yan, Muli Yang and Cheng Deng . . . . . 5004

*QueryAgent: A Reliable and Efficient Reasoning Framework with Environmental Feedback based Self-Correction*
Xiang Huang, Sitao Cheng, Shanshan Huang, Jiayu Shen, Yong Xu, Chaoyun Zhang and Yuzhong Qu . . . . . 5014

*PITA: Prompting Task Interaction for Argumentation Mining*
Yang Sun, Muyi Wang, Jianzhu Bao, Bin Liang, Xiaoyan Zhao, Caihua Yang, Min Yang and Ruifeng Xu . . . . . 5036

*Shifting Attention to Relevance: Towards the Predictive Uncertainty Quantification of Free-Form Large Language Models*
Jinhao Duan, Hao Cheng, Shiqi Wang, Alex Zavalny, Chenan Wang, Renjing Xu, Bhavya Kailkhura and Kaidi Xu . . . . . 5050

*Babel-ImageNet: Massively Multilingual Evaluation of Vision-and-Language Representations*
Gregor Geigle, Radu Timofte and Goran Glavaš . . . . . 5064

*Estimating Agreement by Chance for Sequence Annotation*
Diya Li, Carolyn Rose, Ao Yuan and Chunxiao Zhou . . . . . 5085

*Are Emergent Abilities in Large Language Models just In-Context Learning?*
Sheng Lu, Irina Bigoulaeva, Rachneet Singh Sachdeva, Harish Tayyar Madabushi and Iryna Gurevych . . . . . 5098

*WaveCoder: Widespread And Versatile Enhancement For Code Large Language Models By Instruction Tuning*
Zhaojian Yu, Xin Zhang, Ning Shang, Yangyu Huang, Can Xu, Yishujie Zhao, Wenxiang Hu and Qiufeng Yin . . . . . 5140

*Eliciting Better Multilingual Structured Reasoning from LLMs through Code*
Bryan Li, Tamer Alkhouli, Daniele Bonadiman, Nikolaos Pappas and Saab Mansour . . . . . 5154

*OLIVE: Object Level In-Context Visual Embeddings*
Timothy Ossowski and Junjie Hu . . . . . 5170

*Quantifying Uncertainty in Answers from any Language Model and Enhancing their Trustworthiness*
Jiuhai Chen and Jonas Mueller . . . . . 5186

*Marathon: A Race Through the Realm of Long Context with Large Language Models*
Lei Zhang, Yunshui Li, Ziqiang Liu, Jiaxi Yang, Junhao Liu, Longze Chen, Run Luo and Min Yang . . . . . 5201

*Beyond Scaling: Predicting Patent Approval with Domain-specific Fine-grained Claim Dependency Graph*
Xiaochen Kev Gao, Feng Yao, Kewen Zhao, Beilei He, Animesh Kumar, Vish Krishnan and Jingbo Shang . . . . . 5218

*PCAD: Towards ASR-Robust Spoken Language Understanding via Prototype Calibration and Asymmetric Decoupling*
Xianwei Zhuang, Xuxin Cheng, Liming Liang, Yuxin Xie, Zhichang Wang, Zhiqi Huang and Yuexian Zou . . . . . 5235

*Rethinking the Multimodal Correlation of Multimodal Sequential Learning via Generalizable Attentional Results Alignment*
Tao Jin, Wang Lin, Ye Wang, Linjun Li, Xize Cheng and Zhou Zhao . . . . . 5247

*UHGEval: Benchmarking the Hallucination of Chinese Large Language Models via Unconstrained Generation*
Xun Liang, Shichao Song, Simin Niu, Zhiyu li, Feiyu Xiong, Bo Tang, Yezhaohui Wang, Dawei He, Cheng Peng, Zhonghao Wang and Haiying Deng . . . . . 5266

*PreFLMR: Scaling Up Fine-Grained Late-Interaction Multi-modal Retrievers*
Weizhe Lin, Jingbiao Mei, Jinghong Chen and Bill Byrne . . . . . 5294

*Triple-Encoders: Representations That Fire Together, Wire Together*
Justus-Jonas Erker, Florian Mai, Nils Reimers, Gerasimos Spanakis and Iryna Gurevych . . . 5317

*Improving Hateful Meme Detection through Retrieval-Guided Contrastive Learning*
Jingbiao Mei, Jinghong Chen, Weizhe Lin, Bill Byrne and Marcus Tomalin . . . . . 5333

*Agent-Pro: Learning to Evolve via Policy-Level Reflection and Optimization*
Wenqi Zhang, Ke Tang, Hai Wu, Mengna Wang, Yongliang Shen, Guiyang Hou, Zeqi Tan, Peng Li, Yueting Zhuang and Weiming Lu . . . . . 5348

*Your Transformer is Secretly Linear*
Anton Razzhigaev, Matvey Mikhalchuk, Elizaveta Goncharova, Nikolai Gerasimenko, Ivan Oseledets, Denis Dimitrov and Andrey Kuznetsov . . . . . 5376

*Noise Correction on Subjective Datasets*
Uthman Jinadu and Yi Ding . . . . . 5385

*Generative Explore-Exploit: Training-free Optimization of Generative Recommender Systems using LLM Optimizers*
Lütfi Kerem Senel, Besnik Fetahu, Davis Yoshida, Zhiyu Chen, Giuseppe Castellucci, Nikhita Vedula, Jason Ingyu Choi and Shervin Malmasi . . . . . 5396

*Instruction-tuned Language Models are Better Knowledge Learners*
Zhengbao Jiang, Zhiqing Sun, Weijia Shi, Pedro Rodriguez, Chunting Zhou, Graham Neubig, Xi Victoria Lin, Wen-tau Yih and Srini Iyer . . . . . 5421

*What Do Language Models Hear? Probing for Auditory Representations in Language Models*
Jerry Ngo and Yoon Kim . . . . . 5435

*Threads of Subtlety: Detecting Machine-Generated Texts Through Discourse Motifs*
Zae Myung Kim, Kwang Hee Lee, Preston Zhu, Vipul Raheja and Dongyeop Kang . . . . . 5449

*Jailbreak Open-Sourced Large Language Models via Enforced Decoding*
Hangfan Zhang, Zhimeng Guo, Huaisheng Zhu, Bochuan Cao, Lu Lin, Jinyuan Jia, Jinghui Chen and Dinghao Wu . . . . . 5475

*NICE: To Optimize In-Context Examples or Not?*
Pragya Srivastava, Satvik Golechha, Amit Deshpande and Amit Sharma . . . . . 5494

*CodeScope: An Execution-based Multilingual Multitask Multidimensional Benchmark for Evaluating LLMs on Code Understanding and Generation*
Weixiang Yan, Haitian Liu, Yunkun Wang, Yunzhe Li, Qian Chen, Wen Wang, Tingyu Lin, Weishan Zhao, Li Zhu, Hari Sundaram and Shuiguang Deng . . . . . 5511

*Digital Socrates: Evaluating LLMs through Explanation Critiques*
Yuling Gu, Oyvind Tafjord and Peter Clark . . . . . 5559

*SafeDecoding: Defending against Jailbreak Attacks via Safety-Aware Decoding*
Zhangchen Xu, Fengqing Jiang, Luyao Niu, Jinyuan Jia, Bill Yuchen Lin and Radha Poovendran 5587

*Multi-Task Inference: Can Large Language Models Follow Multiple Instructions at Once?*
Guijin Son, SangWon Baek, Sangdae Nam, Ilgyun Jeong and Seungone Kim . . . . . 5606

*Experiential Co-Learning of Software-Developing Agents*
Chen Qian, Yufan Dang, Jiahao Li, Wei Liu, Zihao Xie, YiFei Wang, Weize Chen, Cheng Yang, Xin Cong, Xiaoyin Che, Zhiyuan Liu and Maosong Sun . . . . . 5628

*Learning Geometry-Aware Representations for New Intent Discovery*
Kai Tang, Junbo Zhao, Xiao Ding, Runze Wu, Lei Feng, Gang Chen and Haobo Wang . . . . . 5641

*Speaker Verification in Agent-generated Conversations*
Yizhe Yang, Palakorn Achananuparp, Heyan Huang, Jing Jiang and Ee-Peng Lim . . . . . 5655

*Benchmarking Data Science Agents*
Yuge Zhang, Qiyang Jiang, XingyuHan XingyuHan, Nan Chen, Yuqing Yang and Kan Ren 5677

*Language-Specific Neurons: The Key to Multilingual Capabilities in Large Language Models*
Tianyi Tang, Wenyang Luo, Haoyang Huang, Dongdong Zhang, Xiaolei Wang, Xin Zhao, Furu Wei and Ji-Rong Wen . . . . . 5701

*Forgetting before Learning: Utilizing Parametric Arithmetic for Knowledge Updating in Large Language Models*
Shiwen Ni, Dingwei Chen, Chengming Li, Xiping Hu, Ruifeng Xu and Min Yang . . . . . 5716

*A Deep Dive into the Trade-Offs of Parameter-Efficient Preference Alignment Techniques*
Megh Thakkar, Quentin Fournier, Matthew D Riemer, Pin-Yu Chen, Amal Zouaq, Payel Das and Sarath Chandar . . . . . 5732

*Zero-Shot Cross-Domain Dialogue State Tracking via Dual Low-Rank Adaptation*
Xiang Luo, Zhiwen Tang, Jin Wang and Xuejie Zhang . . . . . 5746

*PRP-Graph: Pairwise Ranking Prompting to LLMs with Graph Aggregation for Effective Text Re-ranking*
Jian Luo, Xuanang Chen, Ben He and Le Sun . . . . . 5766

*RepCodec: A Speech Representation Codec for Speech Tokenization*
Zhichao Huang, Chutong Meng and Tom Ko . . . . . 5777

*GumbelSoft: Diversified Language Model Watermarking via the GumbelMax-trick*
Jiayi Fu, Xuandong Zhao, Ruihan Yang, Yuansen Zhang, Jiangjie Chen and Yanghua Xiao . 5791

*Event-Radar: Event-driven Multi-View Learning for Multimodal Fake News Detection*
Zihan Ma, Minnan Luo, Hao Guo, Zhi Zeng, Yiran Hao and Xiang Zhao . . . . . 5809

*Fine-Grained Modeling of Narrative Context: A Coherence Perspective via Retrospective Questions*
Liyan Xu, Jiangnan Li, Mo Yu and Jie Zhou . . . . . 5822

*Stealthy Attack on Large Language Model based Recommendation*
Jinghao Zhang, Yuting Liu, Qiang Liu, Shu Wu, Guibing Guo and Liang Wang . . . . . 5839

*Multi-Dimensional Optimization for Text Summarization via Reinforcement Learning*
Sangwon Ryu, Heejin Do, Yunsu Kim, Gary Lee and Jungseul Ok . . . . . 5858

*Masked Thought: Simply Masking Partial Reasoning Steps Can Improve Mathematical Reasoning Learning of Language Models*
Changyu Chen, Xiting Wang, Ting-En Lin, Ang Lv, Yuchuan Wu, Xin Gao, Ji-Rong Wen, Rui Yan and Yongbin Li . . . . . 5872

*SEER: Facilitating Structured Reasoning and Explanation via Reinforcement Learning*
Guoxin Chen, Kexin Tang, Chao Yang, Fuying Ye, Yu Qiao and Yiming Qian . . . . . 5901

*Towards Robust and Generalized Parameter-Efficient Fine-Tuning for Noisy Label Learning*
Yeachan Kim, Junho Kim and SangKeun Lee . . . . . 5922

*SparseFlow: Accelerating Transformers by Sparsifying Information Flows*
Yeachan Kim and SangKeun Lee . . . . . 5937

*ProtT3: Protein-to-Text Generation for Text-based Protein Understanding*
Zhiyuan Liu, An Zhang, Hao Fei, Enzhi Zhang, Xiang Wang, Kenji Kawaguchi and Tat-Seng Chua . . . . . 5949

*KIEval: A Knowledge-grounded Interactive Evaluation Framework for Large Language Models*
Zhuohao Yu, Chang Gao, Wenjin Yao, Yidong Wang, Wei Ye, Jindong Wang, Xing Xie, Yue Zhang and Shikun Zhang . . . . . 5967

*EmoBench: Evaluating the Emotional Intelligence of Large Language Models*
Sahand Sabour, Siyang Liu, Zheyuan Zhang, June M. Liu, Jinfeng Zhou, Alvionna Shiergetya Sunaryo, Tatia M.c. Lee, Rada Mihalcea and Minlie Huang . . . . . 5986

*Are AI-Generated Text Detectors Robust to Adversarial Perturbations?*
Guanhua Huang, Yuchen Zhang, Zhe Li, Yongjian You, Mingze Wang and Zhouwang Yang 6005

*FinTextQA: A Dataset for Long-form Financial Question Answering*
Jian Chen, Peilin Zhou, Yining Hua, Loh Ying Xin, Kehui Chen, Ziyuan Li, Bing Zhu and Junwei Liang . . . . . 6025

*On Measuring Faithfulness or Self-consistency of Natural Language Explanations*
Letitia Parcalabescu and Anette Frank . . . . . 6048

*Learning or Self-aligning? Rethinking Instruction Fine-tuning*
Mengjie Ren, Boxi Cao, Hongyu Lin, Cao Liu, Xianpei Han, Ke Zeng, Wan Guanglu, Xunliang Cai and Le Sun . . . . . 6090

*Rethinking the Bounds of LLM Reasoning: Are Multi-Agent Discussions the Key?*
Qineng Wang, Zihao Wang, Ying Su, Hanghang Tong and Yangqiu Song . . . . . 6106

*Soft Knowledge Prompt: Help External Knowledge Become a Better Teacher to Instruct LLM in Knowledge-based VQA*
Qunbo Wang, Ruyi Ji, Tianhao Peng, Wenjun Wu, Zechao Li and Jing Liu . . . . . 6132

*TasTe: Teaching Large Language Models to Translate through Self-Reflection*
Yutong Wang, Jiali Zeng, Xuebo Liu, Fandong Meng, Jie Zhou and Min Zhang . . . . . 6144

*Not All Experts are Equal: Efficient Expert Pruning and Skipping for Mixture-of-Experts Large Language Models*
Xudong Lu, Qi Liu, Yuhui Xu, Aojun Zhou, Siyuan Huang, Bo Zhang, Junchi Yan and Hongsheng Li . . . . . 6159

*UNIMO-G: Unified Image Generation through Multimodal Conditional Diffusion*
Wei Li, Xue Xu, Jiachen Liu and Xinyan Xiao . . . . . 6173

*The Fine-Tuning Paradox: Boosting Translation Quality Without Sacrificing LLM Abilities*
David Stap, Eva Hasler, Bill Byrne, Christof Monz and Ke Tran . . . . . 6189

*Blinded by Generated Contexts: How Language Models Merge Generated and Retrieved Contexts When Knowledge Conflicts?*
Hexiang Tan, Fei Sun, Wanli Yang, Yuanzhuo Wang, Qi Cao and Xueqi Cheng . . . . . 6207

*Unveiling Linguistic Regions in Large Language Models*
Zhihao Zhang, Jun Zhao, Qi Zhang, Tao Gui and Xuanjing Huang . . . . . 6228

*Text-to-Song: Towards Controllable Music Generation Incorporating Vocal and Accompaniment*
Zhiqing Hong, Rongjie Huang, Xize Cheng, Yongqi Wang, Ruiqi Li, Fuming You, Zhou Zhao and Zhimeng Zhang . . . . . 6248

*FastFiD: Improve Inference Efficiency of Open Domain Question Answering via Sentence Selection*
Yufei Huang, Xu Han and Maosong Sun . . . . . 6262

*Discursive Socratic Questioning: Evaluating the Faithfulness of Language Models' Understanding of Discourse Relations*
Yisong Miao, Hongfu Liu, Wenqiang Lei, Nancy F. Chen and Min-Yen Kan . . . . . 6277

*An Open Multilingual System for Scoring Readability of Wikipedia*
Mykola Trokhymovych, Indira Sen and Martin Gerlach . . . . . 6296

*Unlearning Traces the Influential Training Data of Language Models*
Masaru Isonuma and Ivan Titov . . . . . 6312

*Exploring Alignment in Shared Cross-lingual Spaces*
Basel Mousi, Nadir Durrani, Fahim Dalvi, Majd Hawasly and Ahmed Abdelali . . . . . 6326

*Not All Countries Celebrate Thanksgiving: On the Cultural Dominance in Large Language Models*
Wenxuan Wang, Wenxiang Jiao, Jingyuan Huang, Ruyi Dai, Jen-tse Huang, Zhaopeng Tu and Michael Lyu . . . . . 6349

*Self-Evolving GPT: A Lifelong Autonomous Experiential Learner*
Jinglong Gao, Xiao Ding, Yiming Cui, Jianbai Zhao, Hepeng Wang, Ting Liu and Bing Qin 6385

*WRP: Weight Recover Prune for Structured Sparsity*
Zhendong Tan, Xingjun Zhang and Zheng Wei . . . . . 6433

*Error-preserving Automatic Speech Recognition of Young English Learners' Language*
Janick Michot, Manuela Hürlimann, Jan Milan Deriu, Luzia Sauer, Katsiaryna Mlynchyk and Mark Cieliebak . . . . . 6444

*DiFiNet: Boundary-Aware Semantic Differentiation and Filtration Network for Nested Named Entity Recognition*
Yuxiang Cai, Qiao Liu, Yanglei Gan, Run Lin, Changlin Li, Xueyi Liu, Da Luo and JiayeYang JiayeYang . . . . . 6455

*Legal Case Retrieval: A Survey of the State of the Art*
Yi Feng, Chuanyi Li and Vincent Ng . . . . . 6472

*Benchmarking and Improving Compositional Generalization of Multi-aspect Controllable Text Generation*
Tianqi Zhong, Zhaoyi Li, Quan Wang, Linqi Song, Ying Wei, Defu Lian and Zhendong Mao 6486

*LLaMA Pro: Progressive LLaMA with Block Expansion*
Chengyue Wu, Yukang Gan, Yixiao Ge, Zeyu Lu, Jiahao Wang, Ye Feng, Ying Shan and Ping Luo . . . . . 6518

*Generating Contrastive Narratives Using the Brownian Bridge Process for Narrative Coherence Learning*
Feiteng Mu and Wenjie Li . . . . . 6538

*A Causal Approach for Counterfactual Reasoning in Narratives*
Feiteng Mu and Wenjie Li . . . . . 6556

*SIP: Injecting a Structural Inductive Bias into a Seq2Seq Model by Simulation*
Matthias Lindemann, Alexander Koller and Ivan Titov . . . . . 6570

*The Hidden Space of Transformer Language Adapters*
Jesujoba Oluwadara Alabi, Marius Mosbach, Matan Eyal, Dietrich Klakow and Mor Geva . 6588

*A Ship of Theseus: Curious Cases of Paraphrasing in LLM-Generated Texts*
Nafis Irtiza Tripto, Saranya Venkatraman, Dominik Macko, Robert Moro, Ivan Srba, Adaku Uchendu, Thai Le and Dongwon Lee . . . . . 6608

*Advancing Large Language Models to Capture Varied Speaking Styles and Respond Properly in Spoken Conversations*
Guan-Ting Lin, Cheng-Han Chiang and Hung-yi Lee . . . . . 6626

*RetinaQA: A Robust Knowledge Base Question Answering Model for both Answerable and Unanswerable Questions*
Prayushi Faldu, Indrajit Bhattacharya and Mausam . . . . . 6643

*GroundingGPT: Language Enhanced Multi-modal Grounding Model*
Zhaowei Li, Qi Xu, Dong Zhang, Hang Song, YiQing Cai, Qi Qi, Ran Zhou, Junting Pan, Zefeng Li, Vu Van Tu, Zhida Huang and Tao Wang . . . . . 6657

*Automated Justification Production for Claim Veracity in Fact Checking: A Survey on Architectures and Approaches*
Islam Eldifrawi, Shengrui Wang and Amine Trabelsi . . . . . 6679

*Decoupled Vocabulary Learning Enables Zero-Shot Translation from Unseen Languages*
Carlos Mullov, Quan Pham and Alexander Waibel . . . . . 6693

*SwapMoE: Serving Off-the-shelf MoE-based Large Language Models with Tunable Memory Budget*
Rui Kong, Yuanchun Li, Qingtian Feng, Weijun Wang, Xiaozhou Ye, Ye Ouyang, Linghe Kong and Yunxin Liu . . . . . 6710

*PixT3: Pixel-based Table-To-Text Generation*
Iñigo Alonso, Eneko Agirre and Mirella Lapata . . . . . 6721

*Narrowing the Knowledge Evaluation Gap: Open-Domain Question Answering with Multi-Granularity Answers*
Gal Yona, Roee Aharoni and Mor Geva . . . . . 6737

*TAMS: Translation-Assisted Morphological Segmentation*
Enora Rice, Ali Marashian, Luke Gessler, Alexis Palmer and Katharina Von Der Wense . . . . 6752

*XCodeEval: An Execution-based Large Scale Multilingual Multitask Benchmark for Code Understanding, Generation, Translation and Retrieval*
Mohammad Abdullah Matin Khan, M Saiful Bari, Do Xuan Long, Weishi Wang, Md Rizwan Parvez and Shafiq Joty . . . . 6766

*ProxyQA: An Alternative Framework for Evaluating Long-Form Text Generation with Large Language Models*
Haochen Tan, Zhijiang Guo, Zhan Shi, Lu Xu, Zhili Liu, Yunlong Feng, Xiaoguang Li, Yasheng Wang, Lifeng Shang, Qun Liu and Linqi Song . . . . 6806

*A Glitch in the Matrix? Locating and Detecting Language Model Grounding with Fakepedia*
Giovanni Monea, Maxime Peyrard, Martin Josifoski, Vishrav Chaudhary, Jason Eisner, Emre Kiciman, Hamid Palangi, Barun Patra and Robert West . . . . 6828

*Muffin or Chihuahua? Challenging Multimodal Large Language Models with Multipanel VQA*
Yue Fan, Jing Gu, Kaiwen Zhou, Qianqi Yan, Shan Jiang, Ching-Chen Kuo, Yang Zhao, Xinze Guan and Xin Eric Wang . . . . 6845

*WebVoyager: Building an End-to-End Web Agent with Large Multimodal Models*
Hongliang He, Wenlin Yao, Kaixin Ma, Wenhao Yu, Yong Dai, Hongming Zhang, Zhenzhong Lan and Dong Yu . . . . 6864

*Translation-based Lexicalization Generation and Lexical Gap Detection: Application to Kinship Terms*
Senyu Li, Bradley Hauer, Ning Shi and Grzegorz Kondrak . . . . 6891

*Leveraging Machine-Generated Rationales to Facilitate Social Meaning Detection in Conversations*
Ritam Dutt, Zhen Wu, Jiaxin Shi, Divyanshu Sheth, Prakhar Gupta and Carolyn Rose . . . . 6901

*Robust Frame-Semantic Models with Lexical Unit Trees and Negative Samples*
Jacob Devasier, Yogesh Gurjar and Chengkai Li . . . . 6930

*Harnessing the Power of Large Language Models for Natural Language to First-Order Logic Translation*
Yuan Yang, Siheng Xiong, Ali Payani, Ehsan Shareghi and Faramarz Fekri . . . . 6942

*Lightweight reranking for language model generations*
Siddhartha Jain, Xiaofei Ma, Anoop Deoras and Bing Xiang . . . . 6960

*ARIES: A Corpus of Scientific Paper Edits Made in Response to Peer Reviews*
Mike D'Arcy, Alexis Ross, Erin Bransom, Bailey Kuehl, Jonathan Bragg, Tom Hope and Doug Downey . . . . 6985

*The Unreasonable Effectiveness of Easy Training Data for Hard Tasks*
Peter Hase, Mohit Bansal, Peter Clark and Sarah Wiegreffe . . . . 7002

*PLUG: Leveraging Pivot Language in Cross-Lingual Instruction Tuning*
Zhihan Zhang, Dong-Ho Lee, Yuwei Fang, Wenhao Yu, Mengzhao Jia, Meng Jiang and Francesco Barbieri . . . . 7025

*MIDGARD: Self-Consistency Using Minimum Description Length for Structured Commonsense Reasoning*
Inderjeet Jayakumar Nair and Lu Wang . . . . 7047

*ReConcile: Round-Table Conference Improves Reasoning via Consensus among Diverse LLMs*
Justin Chen, Swarnadeep Saha and Mohit Bansal . . . . . 7066

*Mirror: Multiple-perspective Self-Reflection Method for Knowledge-rich Reasoning*
Hanqi Yan, Qinglin Zhu, Xinyu Wang, Lin Gui and Yulan He . . . . . 7086

*Where Do People Tell Stories Online? Story Detection Across Online Communities*
Maria Antoniak, Joel Mire, Maarten Sap, Elliott Ash and Andrew Piper . . . . . 7104

*Large Language Models Are No Longer Shallow Parsers*
Yuanhe Tian, Fei Xia and Yan Song . . . . . 7131

*Dialogue Summarization with Mixture of Experts based on Large Language Models*
Yuanhe Tian, Fei Xia and Yan Song . . . . . 7143

*ChiMed-GPT: A Chinese Medical Large Language Model with Full Training Regime and Better Alignment to Human Preferences*
Yuanhe Tian, Ruyi Gan, Yan Song, Jiaxing Zhang and Yongdong Zhang . . . . . 7156

*An Investigation of Neuron Activation as a Unified Lens to Explain Chain-of-Thought Eliciting Arithmetic Reasoning of LLMs*
Daking Rai and Ziyu Yao . . . . . 7174

*Leveraging Large Language Models for Learning Complex Legal Concepts through Storytelling*
Hang Jiang, Xiajie Zhang, Robert Mahari, Daniel Kessler, Eric Ma, Tal August, Irene Li, Alex Pentland, Yoon Kim, Deb Roy and Jad Kabbara . . . . . 7194

*Intrinsic Task-based Evaluation for Referring Expression Generation*
Guanyi Chen, Fahime Same and Kees Van Deemter . . . . . 7220

*From Moments to Milestones: Incremental Timeline Summarization Leveraging Large Language Models*
Qisheng Hu, Geonsik Moon and Hwee Tou Ng . . . . . 7232

*End-to-end Learning of Logical Rules for Enhancing Document-level Relation Extraction*
Kunxun Qi, Jianfeng Du and Hai Wan . . . . . 7247

*Can We Achieve High-quality Direct Speech-to-Speech Translation without Parallel Speech Data?*
Qingkai Fang, Shaolei Zhang, Zhengrui Ma, Min Zhang and Yang Feng . . . . . 7264

*Enhancing EEG-to-Text Decoding through Transferable Representations from Pre-trained Contrastive EEG-Text Masked Autoencoder*
Jiaqi Wang, Zhenxi Song, Zhengyu Ma, Xipeng Qiu, Min Zhang and Zhiguo Zhang . . . . . 7278

*CQIL: Inference Latency Optimization with Concurrent Computation of Quasi-Independent Layers*
Longwei Zou, Qingyang Wang, Han Zhao, Jiangangkong Jiangangkong, YI Yang and Yangdong Deng . . . . . 7293

*Prompt Optimization via Adversarial In-Context Learning*
Do Xuan Long, Yiran Zhao, Hannah Brown, Yuxi Xie, James Xu Zhao, Nancy F. Chen, Kenji Kawaguchi, Michael Shieh and Junxian He . . . . . 7308

*StreamVoice: Streamable Context-Aware Language Modeling for Real-time Zero-Shot Voice Conversion*
Zhichao Wang, Yuanzhe Chen, Xinsheng Wang, Lei Xie and Yuping Wang . . . . . 7328

*Generate-then-Ground in Retrieval-Augmented Generation for Multi-hop Question Answering*
Zhengliang Shi, Shuo Zhang, Weiwei Sun, Shen Gao, Pengjie Ren, Zhumin Chen and Zhaochun Ren . . . . . 7339

*Multimodal Contextualized Semantic Parsing from Speech*
Jordan Voas, David Harwath and Ray Mooney . . . . . 7354

*LaMP: When Large Language Models Meet Personalization*
Alireza Salemi, Sheshera Mysore, Michael Bendersky and Hamed Zamani . . . . . 7370

*AboutMe: Using Self-Descriptions in Webpages to Document the Effects of English Pretraining Data Filters*
Li Lucy, Suchin Gururangan, Luca Soldaini, Emma Strubell, David Bamman, Lauren Klein and Jesse Dodge . . . . . 7393

*MT-Bench-101: A Fine-Grained Benchmark for Evaluating Large Language Models in Multi-Turn Dialogues*
Ge Bai, Jie Liu, Xingyuan Bu, Yancheng He, Jiaheng Liu, Zhanhui Zhou, Zhuoran Lin, Wenbo Su, Tiezheng Ge, Bo Zheng and Wanli Ouyang . . . . . 7421

*EFSA: Towards Event-Level Financial Sentiment Analysis*
Tianyu Chen, Yiming Zhang, Guoxin Yu, Dapeng Zhang, Li Zeng, Qing He and Xiang Ao . 7455

*What Evidence Do Language Models Find Convincing?*
Alexander Wan, Eric Wallace and Dan Klein . . . . . 7468

*Advancement in Graph Understanding: A Multimodal Benchmark and Fine-Tuning of Vision-Language Models*
Qihang Ai, Jiafan Li, Jincheng Dai, Jianwu Zhou, Lemao Liu, Haiyun Jiang and Shuming Shi 7485

*LangBridge: Multilingual Reasoning Without Multilingual Supervision*
Dongkeun Yoon, Joel Jang, Sungdong Kim, Seungone Kim, Sheikh Shafayat and Minjoon Seo 7502

*Can LLMs Reason with Rules? Logic Scaffolding for Stress-Testing and Improving LLMs*
Siyuan Wang, Zhongyu Wei, Yejin Choi and Xiang Ren . . . . . 7523

*SEGO: Sequential Subgoal Optimization for Mathematical Problem-Solving*
Xueliang Zhao, Xinting Huang, Wei Bi and Lingpeng Kong . . . . . 7544

*Unlocking the Power of Large Language Models for Entity Alignment*
Xuhui Jiang, Yinghan Shen, Zhichao Shi, Chengjin Xu, Wei Li, Zixuan Li, Jian Guo, Huawei Shen and Yuanzhuo Wang . . . . . 7566

*Trial and Error: Exploration-Based Trajectory Optimization of LLM Agents*
Yifan Song, Da Yin, Xiang Yue, Jie Huang, Sujian Li and Bill Yuchen Lin . . . . . 7584

*ReFT: Reasoning with Reinforced Fine-Tuning*
Luong Quoc Trung, Xinbo Zhang, Zhanming Jie, Peng Sun, Xiaoran Jin and Hang Li . . . . . 7601

*Cognitive Visual-Language Mapper: Advancing Multimodal Comprehension with Enhanced Visual Knowledge Alignment*
Yunxin li, Xinyu Chen, Baotian Hu, Haoyuan Shi and Min Zhang . . . . . 7615

*FreeCtrl: Constructing Control Centers with Feedforward Layers for Learning-Free Controllable Text Generation*
Zijian Feng, Hanzhang Zhou, Kezhi Mao and Zixiao Zhu . . . . . 7627

*HD-Eval: Aligning Large Language Model Evaluators Through Hierarchical Criteria Decomposition*
Yuxuan Liu, Tianchi Yang, Shaohan Huang, Zihan Zhang, Haizhen Huang, Furu Wei, Weiwei Deng, Feng Sun and Qi Zhang . . . . . 7641

*Conundrums in Cross-Prompt Automated Essay Scoring: Making Sense of the State of the Art*
Shengjie Li and Vincent Ng . . . . . 7661

*Angry Men, Sad Women: Large Language Models Reflect Gendered Stereotypes in Emotion Attribution*
Flor Miriam Plaza-del-Arco, Amanda Cercas Curry, Alba Cercas Curry, Gavin Abercrombie and Dirk Hovy . . . . . 7682

*Label Augmentation for Zero-Shot Hierarchical Text Classification*
Lorenzo Paletto, Valerio Basile and Roberto Esposito . . . . . 7697

*STICKERCONV: Generating Multimodal Empathetic Responses from Scratch*
Yiqun Zhang, Fanheng Kong, Peidong Wang, Shuang Sun, SWangLing SWangLing, Shi Feng, Daling Wang, Yifei Zhang and Kaisong Song . . . . . 7707

*EIT: Enhanced Interactive Transformer*
Tong Zheng, Bei Li, Huiwen Bao, Tong Xiao and JingBo Zhu . . . . . 7734

*MARS: Meaning-Aware Response Scoring for Uncertainty Estimation in Generative LLMs*
Yavuz Faruk Bakman, Duygu Nur Yaldiz, Baturalp Buyukates, Chenyang Tao, Dimitrios Dimitriadis and Salman Avestimehr . . . . . 7752

*EXAMS-V: A Multi-Discipline Multilingual Multimodal Exam Benchmark for Evaluating Vision Language Models*
Rocktim Jyoti Das, Simeon Emilov Hristov, Haonan Li, Dimitar Iliyanov Dimitrov, Ivan Koychev and Preslav Nakov . . . . . 7768

*Order-Agnostic Data Augmentation for Few-Shot Named Entity Recognition*
Huiming Wang, Liying Cheng, Wenxuan Zhang, De Wen Soh and Lidong Bing . . . . . 7792

*Text Embedding Inversion Security for Multilingual Language Models*
Yiyi Chen, Heather Lent and Johannes Bjerva . . . . . 7808

*Large Language Models are Superpositions of All Characters: Attaining Arbitrary Role-play via Self-Alignment*
Keming Lu, Bowen Yu, Chang Zhou and Jingren Zhou . . . . . 7828

*PlatoLM: Teaching LLMs in Multi-Round Dialogue via a User Simulator*
Chuyi Kong, Yaxin Fan, Xiang Wan, Feng Jiang and Benyou Wang . . . . . 7841

*Synthesizing Text-to-SQL Data from Weak and Strong LLMs*
Jiaxi Yang, Binyuan Hui, Min Yang, Jian Yang, Junyang Lin and Chang Zhou . . . . . 7864

*STRUCTSUM Generation for Faster Text Comprehension*
Parag Jain, Andreea Marzoca and Francesco Piccinno . . . . . 7876

*Analysing The Impact of Sequence Composition on Language Model Pre-Training*
Yu Zhao, Yuanbin Qu, Konrad Staniszewski, Szymon Tworkowski, Wei Liu, Piotr Miłoś, Yuxiang Wu and Pasquale Minervini . . . . . 7897

*NACL: A General and Effective KV Cache Eviction Framework for LLM at Inference Time*
Yilong Chen, Guoxia Wang, Junyuan Shang, Shiyao Cui, Zhenyu Zhang, Tingwen Liu, Shuohuan Wang, Yu Sun, Dianhai Yu and Hua Wu . . . . . 7913

*SpikeVoice: High-Quality Text-to-Speech Via Efficient Spiking Neural Network*
Kexin Wang, Jiahong Zhang, Yong Ren, Man Yao, Di Shang, Bo XU and Guoqi Li . . . . . 7927

*Context-aware Difference Distilling for Multi-change Captioning*
Yunbin Tu, Liang Li, Li Su, Zheng-Jun Zha, Chenggang Yan and Qingming Huang . . . . . 7941

*Dataflow-Guided Retrieval Augmentation for Repository-Level Code Completion*
Wei Cheng, Yuhan Wu and Wei Hu . . . . . 7957

*Chain-of-Exemplar: Enhancing Distractor Generation for Multimodal Educational Question Generation*
Haohao Luo, Yang Deng, Ying Shen, See-Kiong Ng and Tat-Seng Chua . . . . . 7978

*LLMEmbed: Rethinking Lightweight LLM's Genuine Function in Text Classification*
ChunLiu ChunLiu, Hongguang Zhang, Kainan Zhao, Xinghai Ju and Lin Yang . . . . . 7994

*LEMON: Reviving Stronger and Smaller LMs from Larger LMs with Linear Parameter Fusion*
Yilong Chen, Junyuan Shang, Zhenyu Zhang, Shiyao Cui, Tingwen Liu, Shuohuan Wang, Yu Sun and Hua Wu . . . . . 8005

*Speech Sense Disambiguation: Tackling Homophone Ambiguity in End-to-End Speech Translation*
Tengfei Yu, Xuebo Liu, Liang Ding, Kehai Chen, Dacheng Tao and Min Zhang . . . . . 8020

*To be Continuous, or to be Discrete, Those are Bits of Questions*
Yiran Wang and Masao Utiyama . . . . . 8036

*Moûsai: Efficient Text-to-Music Diffusion Models*
Flavio Schneider, Ojasv Kamal, Zhijing Jin and Bernhard Schölkopf . . . . . 8050

*PokeMQA: Programmable knowledge editing for Multi-hop Question Answering*
Hengrui Gu, Kaixiong Zhou, Xiaotian Han, Ninghao Liu, Ruobing Wang and Xin Wang . . . 8069

*MemeGuard: An LLM and VLM-based Framework for Advancing Content Moderation via Meme Intervention*
Prince Jha, Raghav Jain, Konika Mandal, Aman Chadha, Sriparna Saha and Pushpak Bhattacharyya . . . . . 8084

*Efficient OCR for Building a Diverse Digital History*
Jacob Carlson, Tom Bryan and Melissa Dell . . . . . 8105

*Acquiring Clean Language Models from Backdoor Poisoned Datasets by Downscaling Frequency Space*
Zongru Wu, Zhuosheng Zhang, Pengzhou Cheng and Gongshen Liu . . . . . 8116

*ANAH: Analytical Annotation of Hallucinations in Large Language Models*
Ziwei Ji, Yuzhe Gu, Wenwei Zhang, Chengqi Lyu, Dahua Lin and Kai Chen . . . . . 8135

*Aligning Large Language Models for Controllable Recommendations*
Wensheng Lu, Jianxun Lian, Wei Zhang, Guanghua Li, Mingyang Zhou, Hao Liao and Xing Xie 8159

*Revealing the Parametric Knowledge of Language Models: A Unified Framework for Attribution Methods*
Haeun Yu, Pepa Atanasova and Isabelle Augenstein . . . . . 8173

*Full Parameter Fine-tuning for Large Language Models with Limited Resources*
Kai Lv, Yuqing Yang, Tengxiao Liu, Qipeng Guo and Xipeng Qiu . . . . . 8187

*$M^3$CoT: A Novel Benchmark for Multi-Domain Multi-step Multi-modal Chain-of-Thought*
Qiguang Chen, Libo Qin, Jin Zhang, Zhi Chen, Xiao Xu and Wanxiang Che . . . . . 8199

*Long Context is Not Long at All: A Prospector of Long-Dependency Data for Large Language Models*
Longze Chen, Ziqiang Liu, Wanwei He, Yinhe Zheng, Hao Sun, Yunshui Li, Run Luo and Min Yang . . . . . 8222

*Label-Synchronous Neural Transducer for E2E Simultaneous Speech Translation*
Keqi Deng and Phil Woodland . . . . . 8235

*Hard Prompts Made Interpretable: Sparse Entropy Regularization for Prompt Tuning with RL*
Yunseon Choi, Sangmin Bae, Seonghyun Ban, Minchan Jeong, Chuheng Zhang, Lei Song, Li Zhao, Jiang Bian and Kee-Eung Kim . . . . . 8252

*A Modular Approach for Multimodal Summarization of TV Shows*
Louis Mahon and Mirella Lapata . . . . . 8272

*Think Twice: Perspective-Taking Improves Large Language Models' Theory-of-Mind Capabilities*
Alex Wilf, Sihyun Shawn Lee, Paul Pu Liang and Louis-Philippe Morency . . . . . 8292

*BizBench: A Quantitative Reasoning Benchmark for Business and Finance*
Michael Krumdick, Rik Koncel-Kedziorski, Viet Dac Lai, Varshini Reddy, Charles Lovering and Chris Tanner . . . . . 8309

*Direct Metric Optimization for Image Captioning through Reward-Weighted Augmented Data Utilization*
Takumi Takada, Yuma Suzuki, Hiroki Takushima, Hayato Tanoue, Haruki Sato, Aiswariya Manoj Kumar, Hiroki Nishihara, Takayuki Hori and Kazuya Ueki . . . . . 8333

*Deciphering Hate: Identifying Hateful Memes and Their Targets*
Eftekhar Hossain, Omar Sharif, Mohammed Moshiul Hoque and Sarah Masud Preum . . . . . 8347

*Inducing Systematicity in Transformers by Attending to Structurally Quantized Embeddings*
Yichen Jiang, Xiang Zhou and Mohit Bansal . . . . . 8360

*Label-Efficient Model Selection for Text Generation*
Shir Ashury Tahan, Ariel Gera, Benjamin Sznajder, Leshem Choshen, Liat Ein-Dor and Eyal Shnarch . . . . . 8384

*Machine Unlearning of Pre-trained Large Language Models*
Jin Yao, Eli Chien, Minxin Du, Xinyao Niu, Tianhao Wang, Zezhou Cheng and Xiang Yue . 8403

*Competition of Mechanisms: Tracing How Language Models Handle Facts and Counterfactuals*
Francesco Ortu, Zhijing Jin, Diego Doimo, Mrinmaya Sachan, Alberto Cazzaniga and Bernhard Schölkopf . . . . . 8420

*FactPICO: Factuality Evaluation for Plain Language Summarization of Medical Evidence*
Sebastian Antony Joseph, Lily Chen, Jan Trienes, Hannah Louisa Göke, Monika Coers, Wei Xu, Byron C Wallace and Junyi Jessy Li . . . . . 8437

*BvSP: Broad-view Soft Prompting for Few-Shot Aspect Sentiment Quad Prediction*
Yinhao Bai, Yalan Xie, Xiaoyi Liu, Yuhua Zhao, Zhixin Han, Mengting Hu, Hang Gao and Renhong Cheng . . . . . 8465

*Safety Alignment in NLP Tasks: Weakly Aligned Summarization as an In-Context Attack*
Yu Fu, Yufei Li, Wen Xiao, Cong Liu and Yue Dong . . . . . 8483

*Speech language models lack important brain-relevant semantics*
Subba Reddy Oota, Emin Çelik, Fatma Deniz and Mariya Toneva . . . . . 8503

*DocLLM: A Layout-Aware Generative Language Model for Multimodal Document Understanding*
Dongsheng Wang, Natraj Raman, Mathieu Sibue, Zhiqiang Ma, Petr Babkin, Simerjot Kaur, Yulong Pei, Armineh Nourbakhsh and Xiaomo Liu . . . . . 8529

*Bypassing LLM Watermarks with Color-Aware Substitutions*
Qilong Wu and Varun Chandrasekaran . . . . . 8549

*Parallel Structures in Pre-training Data Yield In-Context Learning*
Yanda Chen, Chen Zhao, Zhou Yu, Kathleen McKeown and He He . . . . . 8582

*OpenToM: A Comprehensive Benchmark for Evaluating Theory-of-Mind Reasoning Capabilities of Large Language Models*
Hainiu Xu, Runcong Zhao, Lixing Zhu, Jinhua Du and Yulan He . . . . . 8593

*Towards Privacy-Aware Sign Language Translation at Scale*
Phillip Rust, Bowen Shi, Skyler Wang, Necati Cihan Camgoz and Jean Maillard . . . . . 8624

*Arithmetic Control of LLMs for Diverse User Preferences: Directional Preference Alignment with Multi-Objective Rewards*
Haoxiang Wang, Yong Lin, Wei Xiong, Rui Yang, Shizhe Diao, Shuang Qiu, Han Zhao and Tong Zhang . . . . . 8642

*Towards Real-World Writing Assistance: A Chinese Character Checking Benchmark with Faked and Misspelled Characters*
Yinghui Li, Zishan Xu, Shaoshen Chen, Haojing Huang, Yangning Li, Shirong Ma, Yong Jiang, Zhongli Li, Qingyu Zhou, Hai-Tao Zheng and Ying Shen . . . . . 8656

*RAVEL: Evaluating Interpretability Methods on Disentangling Language Model Representations*
Jing Huang, Zhengxuan Wu, Christopher Potts, Mor Geva and Atticus Geiger . . . . . 8669

*Large Language Models as Zero-shot Dialogue State Tracker through Function Calling*
Zekun Li, Zhiyu Chen, Mike Ross, Patrick Huber, Seungwhan Moon, Zhaojiang Lin, Xin Luna Dong, Adithya Sagar, Xifeng Yan and Paul A. Crook . . . . . 8688

*Faithful Chart Summarization with ChaTS-Pi*
Syrine Krichene, Francesco Piccinno, Fangyu Liu and Julian Martin Eisenschlos . . . . . 8705

*Enhancing Dialogue State Tracking Models through LLM-backed User-Agents Simulation*
Cheng Niu, Xingguang Wang, Xuxin Cheng, Juntong Song and Tong Zhang . . . . . 8724

*MetaSumPerceiver: Multimodal Multi-Document Evidence Summarization for Fact-Checking*
Ting-Chih Chen, Chia-Wei Tang and Chris Thomas . . . . . 8742

*KnowCoder: Coding Structured Knowledge into LLMs for Universal Information Extraction*
Zixuan Li, Yutao Zeng, Yuxin Zuo, Weicheng Ren, Wenxuan Liu, Miao Su, Yucan Guo, Yantao Liu, Lixiang Lixiang, Zhilei Hu, Long Bai, Wei Li, Yidan Liu, Pan Yang, Xiaolong Jin, Jiafeng Guo and Xueqi Cheng . . . . . 8758

*ERA-CoT: Improving Chain-of-Thought through Entity Relationship Analysis*
Yanming Liu, Xinyue Peng, Tianyu Du, Jianwei Yin, Weihao Liu and Xuhong Zhang . . . . . 8780

*On the Multi-turn Instruction Following for Conversational Web Agents*
Yang Deng, Xuan Zhang, Wenxuan Zhang, Yifei Yuan, See-Kiong Ng and Tat-Seng Chua . 8795

*Mobile-Bench: An Evaluation Benchmark for LLM-based Mobile Agents*
Shihan Deng, Weikai Xu, Hongda Sun, Wei Liu, Tao Tan, Liujianfeng Liujianfeng, Ang Li, Jian Luan, Bin Wang, Rui Yan and Shuo Shang . . . . . 8813

*MC$^2$: Towards Transparent and Culturally-Aware NLP for Minority Languages in China*
Chen Zhang, Mingxu Tao, Quzhe Huang, Jiuheng Lin, Zhibin Chen and Yansong Feng . . . . 8832

*Decoder-only Streaming Transformer for Simultaneous Translation*
Shoutao Guo, Shaolei Zhang and Yang Feng . . . . . 8851

*Defending Large Language Models Against Jailbreaking Attacks Through Goal Prioritization*
Zhexin Zhang, Junxiao Yang, Pei Ke, Fei Mi, Hongning Wang and Minlie Huang . . . . . 8865

*I am a Strange Dataset: Metalinguistic Tests for Language Models*
Tristan Thrush, Jared Moore, Miguel Monares, Christopher Potts and Douwe Kiela . . . . . 8888

*TruthX: Alleviating Hallucinations by Editing Large Language Models in Truthful Space*
Shaolei Zhang, Tian Yu and Yang Feng . . . . . 8908

*ProtLLM: An Interleaved Protein-Language LLM with Protein-as-Word Pre-Training*
Le Zhuo, Zewen Chi, Minghao Xu, Heyan Huang, Jianan Zhao, Heqi Zheng, Conghui He, Xian-Ling Mao and Wentao Zhang . . . . . 8950

*StreamSpeech: Simultaneous Speech-to-Speech Translation with Multi-task Learning*
Shaolei Zhang, Qingkai Fang, Shoutao Guo, Zhengrui Ma, Min Zhang and Yang Feng . . . . . 8964

*Investigating Multi-Hop Factual Shortcuts in Knowledge Editing of Large Language Models*
Tianjie Ju, Yijin Chen, Xinwei Yuan, Zhuosheng Zhang, Wei Du, Yubin Zheng and Gongshen Liu 8987

*Why Don't Prompt-Based Fairness Metrics Correlate?*
Abdelrahman Zayed, Goncalo Mordido, Ioana Baldini and Sarath Chandar . . . . . 9002

*NaijaHate: Evaluating Hate Speech Detection on Nigerian Twitter Using Representative Data*
Manuel Tonneau, Pedro Vitor Quinta De Castro, Karim Lasri, Ibrahim Sambo Farouq, Lakshmi Subramanian, Victor Orozco-Olvera and Samuel Fraiberger . . . . . 9020

*M$^3$AV: A Multimodal, Multigenre, and Multipurpose Audio-Visual Academic Lecture Dataset*
Zhe Chen, Heyang Liu, Wenyi Yu, Guangzhi Sun, Hongcheng Liu, Ji Wu, Chao Zhang, Yu Wang and Yanfeng Wang . . . . . 9041

*Mitigating Biases for Instruction-following Language Models via Bias Neurons Elimination*
Nakyeong Yang, Taegwan Kang, Stanley Jungkyu Choi, Honglak Lee and Kyomin Jung . . . 9061

*Domain Adaptation for Subjective Induction Questions Answering on Products by Adversarial Disentangled Learning*
Yufeng Zhang, Jianxing Yu, Yanghui Rao, Libin Zheng, Qinliang Su, Huaijie Zhu and Jian Yin 9074

*Revisiting Demonstration Selection Strategies in In-Context Learning*
Keqin Peng, Liang Ding, Yancheng Yuan, Xuebo Liu, Min Zhang, Yuanxin Ouyang and Dacheng Tao . . . . . 9090

*Multimodal Table Understanding*
Mingyu Zheng, Xinwei Feng, Qingyi Si, Qiaoqiao She, Zheng Lin, Wenbin Jiang and Weiping Wang . . . . . 9102

*Ex$^3$: Automatic Novel Writing by Extracting, Excelsior and Expanding*
Huang Lei, Jiaming Guo, Guanhua He, Xishan Zhang, Rui Zhang, Shaohui Peng, Shaoli Liu and Tianshi Chen . . . . . 9125

*Few-shot Transfer Learning for Knowledge Base Question Answering: Fusing Supervised Models with In-Context Learning*
Mayur Patidar, Riya Sawhney, Avinash Kumar Singh, Biswajit Chatterjee, Mausam . and Indrajit Bhattacharya . . . . . 9147

*WatME: Towards Lossless Watermarking Through Lexical Redundancy*
Liang Chen, Yatao Bian, Yang Deng, Deng Cai, Shuaiyi Li, Peilin Zhao and Kam-Fai Wong9166

*Text-like Encoding of Collaborative Information in Large Language Models for Recommendation*
Yang Zhang, Keqin Bao, Ming Yan, Wenjie Wang, Fuli Feng and Xiangnan He . . . . . 9181

*MM-SAP: A Comprehensive Benchmark for Assessing Self-Awareness of Multimodal Large Language Models in Perception*
Yuhao Wang, Yusheng Liao, Heyang Liu, Hongcheng Liu, Yanfeng Wang and Yu Wang . . . 9192

*Focus on Your Question! Interpreting and Mitigating Toxic CoT Problems in Commonsense Reasoning*
Jiachun Li, Pengfei Cao, Chenhao Wang, Zhuoran Jin, Yubo Chen, Daojian Zeng, Kang Liu and Jun Zhao . . . . . 9206

*Multi-Aspect Controllable Text Generation with Disentangled Counterfactual Augmentation*
Yi Liu, Xiangyu Liu, Xiangrong Zhu and Wei Hu . . . . . 9231

*Reward-based Input Construction for Cross-document Relation Extraction*
Byeonghu Na, Suhyeon Jo, Yeongmin Kim and Il-chul Moon . . . . . 9254

*Hyperspherical Multi-Prototype with Optimal Transport for Event Argument Extraction*
Guangjun Zhang, Hu Zhang, YuJie Wang, Ru Li, Hongye Tan and Jiye Liang . . . . . 9271

*Understanding Retrieval Robustness for Retrieval-augmented Image Captioning*
Wenyan Li, Jiaang Li, Rita Ramos, Raphael Tang and Desmond Elliott . . . . . 9285

*Semi-Supervised Spoken Language Glossification*
Huijie Yao, Wengang Zhou, Hao Zhou and Houqiang Li . . . . . 9300

*SeeClick: Harnessing GUI Grounding for Advanced Visual GUI Agents*
Kanzhi Cheng, Qiushi Sun, Yougang Chu, Fangzhi Xu, Li YanTao, Jianbing Zhang and Zhiyong Wu . . . . . 9313

*InterrogateLLM: Zero-Resource Hallucination Detection in LLM-Generated Answers*
Yakir Yehuda, Itzik Malkiel, Oren Barkan, Jonathan Weill, Royi Ronen and Noam Koenigstein 9333

*F-Eval: Asssessing Fundamental Abilities with Refined Evaluation Methods*
Yu Sun, Keyuchen Keyuchen, Shujie Wang, Peiji Li, Qipeng Guo, Hang Yan, Xipeng Qiu, Xuanjing Huang and Dahua Lin . . . . . 9348

*Comparing Inferential Strategies of Humans and Large Language Models in Deductive Reasoning*
Philipp Mondorf and Barbara Plank . . . . . 9370

*Whose Preferences? Differences in Fairness Preferences and Their Impact on the Fairness of AI Utilizing Human Feedback*
Maria Emilia Agis Lerner, Florian E. Dorner, Elliott Ash and Naman Goel . . . . . 9403

*Math-Shepherd: Verify and Reinforce LLMs Step-by-step without Human Annotations*
Peiyi Wang, Lei Li, Zhihong Shao, Runxin Xu, Damai Dai, Yifei Li, Deli Chen, Yu Wu and Zhifang Sui . . . . . 9426

*Large Language Models are not Fair Evaluators*
Peiyi Wang, Lei Li, Liang Chen, Zefan Cai, Dawei Zhu, Binghuai Lin, Yunbo Cao, Lingpeng Kong, Qi Liu, Tianyu Liu and Zhifang Sui . . . . . 9440

*Improving Large Language Models in Event Relation Logical Prediction*
Meiqi Chen, Yubo Ma, Kaitao Song, Yixin Cao, Yan Zhang and Dongsheng Li . . . . . 9451

*Synchronized Video Storytelling: Generating Video Narrations with Structured Storyline*
Dingyi Yang, Chunru Zhan, Ziheng Wang, Biao Wang, Tiezheng Ge, Bo Zheng and Qin Jin 9479

*Fine-Grained Image-Text Alignment in Medical Imaging Enables Explainable Cyclic Image-Report Generation*
Wenting Chen, Linlin Shen, Jingyang Lin, Jiebo Luo, Xiang Li and Yixuan Yuan . . . . . 9494

*T-Eval: Evaluating the Tool Utilization Capability of Large Language Models Step by Step*
Zehui Chen, Weihua Du, Wenwei Zhang, Kuikun Liu, Jiangning Liu, Miao Zheng, Jingming Zhuo, Songyang Zhang, Dahua Lin, Kai Chen and Feng Zhao . . . . . 9510

*Are LLM-based Evaluators Confusing NLG Quality Criteria?*
Xinyu Hu, Mingqi Gao, Sen Hu, Yang Zhang, Yicheng Chen, Teng XU and Xiaojun Wan . . 9530

*Synergistic Interplay between Search and Large Language Models for Information Retrieval*
Jiazhan Feng, Chongyang Tao, Xiubo Geng, Tao Shen, Can Xu, Guodong Long, Dongyan Zhao and Daxin Jiang . . . . . 9571

*Linear Transformers with Learnable Kernel Functions are Better In-Context Models*
Yaroslav Aksenov, Nikita Balagansky, Sofia Maria Lo Cicero Vaina, Boris Shaposhnikov, Alexey Gorbatovski and Daniil Gavrilov . . . . . 9584

*Temperature-scaling surprisal estimates improve fit to human reading times – but does it do so for the "right reasons"?*
Tong Liu, Iza Škrjanec and Vera Demberg . . . . . 9598

*Beyond Recognising Entailment: Formalising Natural Language Inference from an Argumentative Perspective*
Ameer Saadat-Yazdi and Nadin Kökciyan . . . . . 9620

*AnyGPT: Unified Multimodal LLM with Discrete Sequence Modeling*
Jun Zhan, Junqi Dai, Jiasheng Ye, Yunhua Zhou, Dong Zhang, Zhigeng Liu, Xin Zhang, Ruibin Yuan, Ge Zhang, Linyang Li, Hang Yan, Jie Fu, Tao Gui, Tianxiang Sun, Yu-Gang Jiang and Xipeng Qiu . . . . . 9637

*CofiPara: A Coarse-to-fine Paradigm for Multimodal Sarcasm Target Identification with Large Multimodal Models*
Zixin Chen, Hongzhan Lin, Ziyang Luo, Mingfei Cheng, Jing Ma and Guang Chen . . . . . 9663

*Direct Large Language Model Alignment Through Self-Rewarding Contrastive Prompt Distillation*
Aiwei Liu, Haoping Bai, Zhiyun Lu, Xiang Kong, Xiaoming Simon Wang, Jiulong Shan, Meng Cao and Lijie Wen . . . . . 9688

*Diffusion Lens: Interpreting Text Encoders in Text-to-Image Pipelines*
Michael Toker, Hadas Orgad, Mor Ventura, Dana Arad and Yonatan Belinkov . . . . . 9713

*Parrot: Enhancing Multi-Turn Instruction Following for Large Language Models*
Yuchong Sun, Che Liu, Kun Zhou, Jinwen Huang, Ruihua Song, Xin Zhao, Fuzheng Zhang, Di Zhang and Kun Gai . . . . . 9729

*Robust Singing Voice Transcription Serves Synthesis*
Ruiqi Li, Yu Zhang, Yongqi Wang, Zhiqing Hong, Rongjie Huang and Zhou Zhao . . . . . 9751

*VulLibGen: Generating Names of Vulnerability-Affected Packages via a Large Language Model*
Tianyu Chen, Lin Li, ZhuLiuchuan ZhuLiuchuan, Zongyang Li, Xueqing Liu, Guangtai Liang, Qianxiang Wang and Tao Xie . . . . . 9767

*Self-Modifying State Modeling for Simultaneous Machine Translation*
Donglei Yu, Xiaomian Kang, Yuchen Liu, Yu Zhou and Chengqing Zong . . . . . 9781

*MapGPT: Map-Guided Prompting with Adaptive Path Planning for Vision-and-Language Navigation*
Jiaqi Chen, Bingqian Lin, Ran Xu, Zhenhua Chai, Xiaodan Liang and Kwan-Yee K. Wong . 9796

*BadAgent: Inserting and Activating Backdoor Attacks in LLM Agents*
Yifei Wang, Dizhan Xue, Shengjie Zhang and Shengsheng Qian . . . . . 9811

*DetermLR: Augmenting LLM-based Logical Reasoning from Indeterminacy to Determinacy*
Hongda Sun, Weikai Xu, Wei Liu, Jian Luan, Bin Wang, Shuo Shang, Ji-Rong Wen and Rui Yan 9828

*LePaRD: A Large-Scale Dataset of Judicial Citations to Precedent*
Robert Mahari, Dominik Stammbach, Elliott Ash and Alex Pentland . . . . . 9863

*To Generate or to Retrieve? On the Effectiveness of Artificial Contexts for Medical Open-Domain Question Answering*
Giacomo Frisoni, Alessio Cocchieri, Alex Presepi, Gianluca Moro and Zaiqiao Meng . . . . . 9878

*MERA: A Comprehensive LLM Evaluation in Russian*
Alena Fenogenova, Artem Chervyakov, Nikita Martynov, Anastasia Kozlova, Maria Tikhonova, Albina Akhmetgareeva, Anton Emelyanov, Denis Shevelev, Pavel Lebedev, Leonid S Sinev, Ulyana Isaeva, Katerina Kolomeytseva, Daniil Moskovskiy, Elizaveta Goncharova, Nikita Savushkin, Polina Mikhailova, Anastasia Minaeva, Denis Dimitrov, Alexander Panchenko and Sergey Markov . . . . . 9920

*SC2: Towards Enhancing Content Preservation and Style Consistency in Long Text Style Transfer*
Jie Zhao, Ziyu Guan, Cai Xu, Wei Zhao and Yue Jiang . . . . . 9949

*Dodo: Dynamic Contextual Compression for Decoder-only LMs*
Guanghui Qin, Corby Rosset, Ethan C. Chau, Nikhil Rao and Benjamin Van Durme . . . . . 9961

*POMP: Probability-driven Meta-graph Prompter for LLMs in Low-resource Unsupervised Neural Machine Translation*
Shilong Pan, Zhiliang Tian, Liang Ding, Haoqi Zheng, Zhen Huang, Zhihua Wen and Dongsheng Li . . . . . 9976

*NewsBench: A Systematic Evaluation Framework for Assessing Editorial Capabilities of Large Language Models in Chinese Journalism*
Miao Li, Ming-Bin Chen, Bo Tang, ShengbinHou ShengbinHou, Pengyu Wang, Haiying Deng, Zhiyu li, Feiyu Xiong, Keming Mao, Cheng Peng and Yi Luo . . . . . 9993

*MAPO: Advancing Multilingual Reasoning through Multilingual-Alignment-as-Preference Optimization*
Shuaijie She, Wei Zou, Shujian Huang, Wenhao Zhu, Xiang Liu, Xiang Geng and Jiajun Chen 10015

*Enhancing Noise Robustness of Retrieval-Augmented Language Models with Adaptive Adversarial Training*
Feiteng Fang, Yuelin Bai, Shiwen Ni, Min Yang, Xiaojun Chen and Ruifeng Xu ..... 10028

*Predicting Text Preference Via Structured Comparative Reasoning*
Jing Nathan Yan, Tianqi Liu, Justin T Chiu, Jiaming Shen, Zhen Qin, Yue Yu, Charumathi Lakshmanan, Yair Kurzion, Alexander M Rush, Jialu Liu and Michael Bendersky ..... 10040

*CoELM: Construction-Enhanced Language Modeling*
Lvxiaowei Xu, Zhilin Gong, Jianhua Dai, Tianxiang Wang, Ming Cai and Jiawei Peng .... 10061

*Uni-Dubbing: Zero-Shot Speech Synthesis from Visual Articulation*
Songju Lei, Xize Cheng, Mengjiao Lyu, Jianqiao Hu, Jintao Tan, Runlin Liu, Lingyu Xiong, Tao Jin, Xiandong Li and Zhou Zhao ..... 10082

*On the Impact of Calibration Data in Post-training Quantization and Pruning*
Miles Williams and Nikolaos Aletras ..... 10100

*SymKGQA: Few-Shot Knowledge Graph Question Answering via Symbolic Program Generation and Execution*
Prerna Agarwal, Nishant Kumar and Srikanta J. Bedathur ..... 10119

*Meta-Task Prompting Elicits Embeddings from Large Language Models*
Yibin Lei, Di Wu, Tianyi Zhou, Tao Shen, Yu Cao, Chongyang Tao and Andrew Yates .... 10141

*A Sentiment Consolidation Framework for Meta-Review Generation*
Miao Li, Jey Han Lau and Eduard Hovy ..... 10158

*Revisiting Structured Sentiment Analysis as Latent Dependency Graph Parsing*
Chengjie Zhou, Bobo Li, Hao Fei, Fei Li, Chong Teng and Donghong Ji ..... 10178

*OWSM-CTC: An Open Encoder-Only Speech Foundation Model for Speech Recognition, Translation, and Language Identification*
Yifan Peng, Yui Sudo, Muhammad Shakeel and Shinji Watanabe ..... 10192

*Do Large Language Models Latently Perform Multi-Hop Reasoning?*
Sohee Yang, Elena Gribovskaya, Nora Kassner, Mor Geva and Sebastian Riedel ..... 10210

*MuggleMath: Assessing the Impact of Query and Response Augmentation on Math Reasoning*
Chengpeng Li, Zheng Yuan, Hongyi Yuan, Guanting Dong, Keming Lu, Jiancan Wu, Chuanqi Tan, Xiang Wang and Chang Zhou ..... 10230

*Harnessing Toulmin's theory for zero-shot argument explication*
Ankita Gupta, Ethan Zuckerman and Brendan O'Connor ..... 10259

*BinaryAlign: Word Alignment as Binary Sequence Labeling*
Gaetan Lopez Latouche, Marc-André Carbonneau and Benjamin Swanson ..... 10277

*Quantifying the Persona Effect in LLM Simulations*
Tiancheng Hu and Nigel Collier ..... 10289

*Artifacts or Abduction: How Do LLMs Answer Multiple-Choice Questions Without the Question?*
Nishant Balepur, Abhilasha Ravichander and Rachel Rudinger ..... 10308

*Retrieval Augmented Fact Verification by Synthesizing Contrastive Arguments*
Zhenrui Yue, Huimin Zeng, Lanyu Shang, Yifan Liu, Yang Zhang and Dong Wang ..... 10331

*SyllabusQA: A Course Logistics Question Answering Dataset*
Nigel Fernandez, Alexander Scarlatos and Andrew Lan . . . . . 10344

*MindMap: Knowledge Graph Prompting Sparks Graph of Thoughts in Large Language Models*
Yilin Wen, Zifeng Wang and Jimeng Sun . . . . . 10370

*AGB-DE: A Corpus for the Automated Legal Assessment of Clauses in German Consumer Contracts*
Daniel Braun and Florian Matthes . . . . . 10389

*Examining the robustness of LLM evaluation to the distributional assumptions of benchmarks*
Charlotte Siska, Katerina Marazopoulou, Melissa Ailem and James Bono . . . . . 10406

*Re-Tuning: Overcoming the Compositionality Limits of Large Language Models with Recursive Tuning*
Eric Pasewark, Kyle Montgomery, Kefei Duan, Dawn Song and Chenguang Wang . . . . . 10422

*Bridging the Preference Gap between Retrievers and LLMs*
Zixuan Ke, Weize Kong, Cheng Li, Mingyang Zhang, Qiaozhu Mei and Michael Bendersky 10438

*Large Language Models Can Learn Temporal Reasoning*
Siheng Xiong, Ali Payani, Ramana Rao Kompella and Faramarz Fekri . . . . . 10452

*Learning Relational Decomposition of Queries for Question Answering from Tables*
Raphaël Mouravieff, Benjamin Piwowarski and Sylvain Lamprier . . . . . 10471

*Characterizing Similarities and Divergences in Conversational Tones in Humans and LLMs by Sampling with People*
Dun-Ming Huang, Pol Van Rijn, Ilia Sucholutsky, Raja Marjieh and Nori Jacoby . . . . . 10486

*Pareto Optimal Learning for Estimating Large Language Model Errors*
Theodore Zhao, Mu Wei, J. Samuel Preston and Hoifung Poon . . . . . 10513

*Simul-LLM: A Framework for Exploring High-Quality Simultaneous Translation with Large Language Models*
Victor Agostinelli, Max Wild, Matthew Raffel, Kazi Ahmed Asif Fuad and Lizhong Chen 10530

*Defending Against Alignment-Breaking Attacks via Robustly Aligned LLM*
Bochuan Cao, Yuanpu Cao, Lu Lin and Jinghui Chen . . . . . 10542

*Interactive-KBQA: Multi-Turn Interactions for Knowledge Base Question Answering with Large Language Models*
Guanming Xiong, Junwei Bao and Wen Zhao . . . . . 10561

*LLMs in the Imaginarium: Tool Learning through Simulated Trial and Error*
Boshi Wang, Hao Fang, Jason Eisner, Benjamin Van Durme and Yu Su . . . . . 10583

*HyperMoE: Towards Better Mixture of Experts via Transferring Among Experts*
Hao Zhao, Zihan Qiu, Huijia Wu, Zili Wang, Zhaofeng He and Jie Fu . . . . . 10605

*Aligning Large Language Models with Human Preferences through Representation Engineering*
Wenhao Liu, Xiaohua Wang, Muling Wu, Tianlong Li, Changze Lv, Zixuan Ling, Zhu JianHao, Cenyuan Zhang, Xiaoqing Zheng and Xuanjing Huang . . . . . 10619

*CODIS: Benchmarking Context-dependent Visual Comprehension for Multimodal Large Language Models*
Fuwen Luo, Chi Chen, Zihao Wan, Zhaolu Kang, Qidong Yan, Yingjie Li, Xiaolong Wang, Siyu Wang, Ziyue Wang, Xiaoyue Mi, Peng Li, Ning Ma, Maosong Sun and Yang Liu . . . . . 10639

*ARAIDA: Analogical Reasoning-Augmented Interactive Data Annotation*
Chen Huang, Yiping Jin, Ilija Ilievski, Wenqiang Lei and Jiancheng Lv . . . . . 10660

*PolCLIP: A Unified Image-Text Word Sense Disambiguation Model via Generating Multimodal Complementary Representations*
Qihao Yang, Yong Li, Xuelin Wang, Fu Lee Wang and Tianyong Hao . . . . . 10676

*Prompted Aspect Key Point Analysis for Quantitative Review Summarization*
An Quang Tang, Xiuzhen Zhang, Minh Ngoc Dinh and Erik Cambria . . . . . 10691

*Ask Again, Then Fail: Large Language Models' Vacillations in Judgment*
Qiming Xie, Zengzhi Wang, Yi Feng and Rui Xia . . . . . 10709

*CLAMBER: A Benchmark of Identifying and Clarifying Ambiguous Information Needs in Large Language Models*
Tong Zhang, Peixin Qin, Yang Deng, Chen Huang, Wenqiang Lei, Junhong Liu, Dingnan Jin, Hongru Liang and Tat-Seng Chua . . . . . 10746

*Multimodal Reasoning with Multimodal Knowledge Graph*
Junlin Lee, Yequan Wang, Jing Li and Min Zhang . . . . . 10767

*Confidence is not Timeless: Modeling Temporal Validity for Rule-based Temporal Knowledge Graph Forecasting*
Rikui Huang, Wei Wei, Xiaoye Qu, Shengzhe Zhang, Dangyang Chen and Yu Cheng . . . . . 10783

*CARE: A Clue-guided Assistant for CSRs to Read User Manuals*
Weihong Du, Jia Liu, Zujie Wen, Dingnan Jin, Hongru Liang and Wenqiang Lei . . . . . 10795

*Enhancing Numerical Reasoning with the Guidance of Reliable Reasoning Processes*
Dingzirui Wang, Longxu Dou, Xuanliang Zhang, Qingfu Zhu and Wanxiang Che . . . . . 10812

*PAGED: A Benchmark for Procedural Graphs Extraction from Documents*
Weihong Du, Wenrui Liao, Hongru Liang and Wenqiang Lei . . . . . 10829

*Navigating the Shadows: Unveiling Effective Disturbances for Modern AI Content Detectors*
Ying Zhou, Ben He and Le Sun . . . . . 10847

*RAGTruth: A Hallucination Corpus for Developing Trustworthy Retrieval-Augmented Language Models*
Cheng Niu, Yuanhao Wu, Juno Zhu, Siliang Xu, KaShun Shum, Randy Zhong, Juntong Song and Tong Zhang . . . . . 10862

*The Dawn After the Dark: An Empirical Study on Factuality Hallucination in Large Language Models*
Junyi Li, Jie Chen, Ruiyang Ren, Xiaoxue Cheng, Xin Zhao, Jian-Yun Nie and Ji-Rong Wen 10879

*Revisiting Knowledge Distillation for Autoregressive Language Models*
Qihuang Zhong, Liang Ding, Li Shen, Juhua Liu, Bo Du and Dacheng Tao . . . . . 10900

*Continual Learning with Semi-supervised Contrastive Distillation for Incremental Neural Machine Translation*
Yunlong Liang, Fandong Meng, Jiaan Wang, Jinan Xu, Yufeng Chen and Jie Zhou . . . . . 10914

*Make-A-Voice: Revisiting Voice Large Language Models as Scalable Multilingual and Multitask Learners*
Rongjie Huang, Chunlei Zhang, Yongqi Wang, Dongchao Yang, Jinchuan Tian, Zhenhui Ye, Luping Liu, Zehan Wang, Ziyue Jiang, Xuankai Chang, Jiatong Shi, Chao Weng, Zhou Zhao and Dong Yu . . . . . 10929

*Chat Vector: A Simple Approach to Equip LLMs with Instruction Following and Model Alignment in New Languages*
Shih-Cheng Huang, Pin-Zu Li, Yu-chi Hsu, Kuang-Ming Chen, Yu Tung Lin, Shih-Kai Hsiao, Richard Tzong-Han Tsai and Hung-yi Lee . . . . . 10943

*PRP: Propagating Universal Perturbations to Attack Large Language Model Guard-Rails*
Neal Mangaokar, Ashish Hooda, Jihye Choi, Shreyas Chandrashekaran, Kassem Fawaz, Somesh Jha and Atul Prakash . . . . . 10960

*Hide and Seek in Noise Labels: Noise-Robust Collaborative Active Learning with LLMs-Powered Assistance*
Bo Yuan, Yulin Chen, Yin Zhang and Wei Jiang . . . . . 10977

*CLOMO: Counterfactual Logical Modification with Large Language Models*
Yinya Huang, Ruixin Hong, Hongming Zhang, Wei Shao, Zhicheng Yang, Dong Yu, Changshui Zhang, Xiaodan Liang and Linqi Song . . . . . 11012

*Exploring Hybrid Question Answering via Program-based Prompting*
Qi Shi, Han Cui, Haofeng Wang, Qingfu Zhu, Wanxiang Che and Ting Liu . . . . . 11035

*IndicGenBench: A Multilingual Benchmark to Evaluate Generation Capabilities of LLMs on Indic Languages*
Harman Singh, Nitish Gupta, Shikhar Bharadwaj, Dinesh Tewari and Partha Talukdar . . . . 11047

*Simple but Effective Compound Geometric Operations for Temporal Knowledge Graph Completion*
Rui Ying, Mengting Hu, Jianfeng Wu, Yalan Xie, Xiaoyi Liu, Zhunheng Wang, Ming Jiang, Hang Gao, Linlin Zhang and Renhong Cheng . . . . . 11074

*Uncertainty Aware Learning for Language Model Alignment*
Yikun Wang, Rui Zheng, Liang Ding, Qi Zhang, Dahua Lin and Dacheng Tao . . . . . 11087

*Interpretable User Satisfaction Estimation for Conversational Systems with Large Language Models*
Ying-Chun Lin, Jennifer Neville, Jack W Stokes, Longqi Yang, Tara Safavi, Mengting Wan, Scott Counts, Siddharth Suri, Reid Andersen, Xiaofeng Xu, Deepak Gupta, Sujay Kumar Jauhar, Xia Song, Georg Buscher, Saurabh Tiwary, Brent Hecht and Jaime Teevan . . . . . 11100

*Fundamental Capabilities of Large Language Models and their Applications in Domain Scenarios: A Survey*
Jiawei Li, Yizhe Yang, Yu Bai, Xiaofeng Zhou, Yinghao Li, Huashan Sun, Yuhang Liu, Xingpeng Si, Yuhao Ye, Yixiao Wu, , Bin Xu, Ren Bowen, Chong Feng, Yang Gao and Heyan Huang . . . . 11116

*Measuring Political Bias in Large Language Models: What Is Said and How It Is Said*
Yejin Bang, Delong Chen, Nayeon Lee and Pascale Fung . . . . . 11142

*Fortify the Shortest Stave in Attention: Enhancing Context Awareness of Large Language Models for Effective Tool Use*
Yuhan Chen, Ang Lv, Ting-En Lin, Changyu Chen, Yuchuan Wu, Fei Huang, Yongbin Li and Rui Yan . . . . . 11160

*Layer-Condensed KV Cache for Efficient Inference of Large Language Models*
Haoyi Wu and Kewei Tu . . . . . 11175

*Enhancing Multilingual Capabilities of Large Language Models through Self-Distillation from Resource-Rich Languages*
Yuanchi Zhang, Yile Wang, Zijun Liu, Shuo Wang, Xiaolong Wang, Peng Li, Maosong Sun and Yang Liu . . . . . 11189

*Benchmarking Chinese Commonsense Reasoning of LLMs: From Chinese-Specifics to Reasoning-Memorization Correlations*
Jiaxing Sun, Weiquan Huang, Jiang Wu, Chenya Gu, Wei Li, Songyang Zhang, Hang Yan and Conghui He . . . . . 11205

*Browse and Concentrate: Comprehending Multimodal Content via Prior-LLM Context Fusion*
Ziyue Wang, Chi Chen, Yiqi Zhu, Fuwen Luo, Peng Li, Ming Yan, Ji Zhang, Fei Huang, Maosong Sun and Yang Liu . . . . . 11229

*Model Composition for Multimodal Large Language Models*
Chi Chen, Yiyang Du, Zheng Fang, Ziyue Wang, Fuwen Luo, Peng Li, Ming Yan, Ji Zhang, Fei Huang, Maosong Sun and Yang Liu . . . . . 11246

*Draft*
*Verify: Lossless Large Language Model Acceleration via Self-Speculative Decoding*
Jun Zhang, Jue Wang, Huan Li, Lidan Shou, Ke Chen, Gang Chen and Sharad Mehrotra . . 11263

*Soul-Mix: Enhancing Multimodal Machine Translation with Manifold Mixup*
Xuxin Cheng, Ziyu Yao, Yifei Xin, Hao An, Hongxiang Li, Yaowei Li and Yuexian Zou . . 11283

*Measuring Meaning Composition in the Human Brain with Composition Scores from Large Language Models*
Changjiang Gao, Jixing Li, Jiajun Chen and Shujian Huang . . . . . 11295

*MIST: Mutual Information Maximization for Short Text Clustering*
Krissanee Kamthawee, Can Udomcharoenchaikit and Sarana Nutanong . . . . . 11309

*Self-chats from Large Language Models Make Small Emotional Support Chatbot Better*
Zhonghua Zheng, Lizi Liao, Yang Deng, Libo Qin and Liqiang Nie . . . . . 11325

*Improving Conversational Abilities of Quantized Large Language Models via Direct Preference Alignment*
Janghwan Lee, Seongmin Park, Sukjin Hong, Minsoo Kim, Du-Seong Chang and Jungwook Choi 11346

*Complex Reasoning over Logical Queries on Commonsense Knowledge Graphs*
Tianqing Fang, Zeming Chen, Yangqiu Song and Antoine Bosselut . . . . . 11365

*An Expert is Worth One Token: Synergizing Multiple Expert LLMs as Generalist via Expert Token Routing*
Ziwei Chai, Guoyin Wang, Jing Su, Tianjie Zhang, Xuanwen Huang, Xuwu Wang, Jingjing Xu, Jianbo Yuan, Hongxia Yang, Fei Wu and Yang Yang . . . . . 11385

*Learning to Plan and Generate Text with Citations*
Constanza Fierro, Reinald Kim Amplayo, Fantine Huot, Nicola De Cao, Joshua Maynez, Shashi Narayan and Mirella Lapata . . . . . 11397

*Exploring Precision and Recall to assess the quality and diversity of LLMs*
Florian Le Bronnec, Alexandre Verine, Benjamin Negrevergne, Yann Chevaleyre and Alexandre Allauzen . . . . . 11418

*Aligning Large Language Models by On-Policy Self-Judgment*
Sangkyu Lee, Sungdong Kim, Ashkan Yousefpour, Minjoon Seo, Kang Min Yoo and Youngjae Yu . . . . . 11442

*IL-TUR: Benchmark for Indian Legal Text Understanding and Reasoning*
Abhinav Joshi, Shounak Paul, Akshat Sharma, Pawan Goyal, Saptarshi Ghosh and Ashutosh Modi 11460

*JumpCoder: Go Beyond Autoregressive Coder via Online Modification*
Mouxiang Chen, Hao Tian, Zhongxin Liu, Xiaoxue Ren and Jianling Sun . . . . . 11500

*Aya Dataset: An Open-Access Collection for Multilingual Instruction Tuning*
Shivalika Singh, Freddie Vargus, Daniel D'souza, Börje F. Karlsson, Abinaya Mahendiran, Wei-Yin Ko, Herumb Shandilya, Jay Patel, Deividas Mataciunas, Laura O'Mahony, Mike Zhang, Ramith Hettiarachchi, Joseph Wilson, Marina Machado, Luisa Souza Moura, Dominik Krzemiński, Hakimeh Fadaei, Irem Ergun, Ifeoma Okoh, Aisha Alaagib, Oshan Ivantha Mudannayake, Zaid Alyafeai, Vu Minh Chien, Sebastian Ruder, Surya Guthikonda, Emad A. Alghamdi, Sebastian Gehrmann, Niklas Muennighoff, Max Bartolo, Julia Kreutzer, Ahmet Üstün, Marzieh Fadaee and Sara Hooker . . . . 11521

*Language Models can Exploit Cross-Task In-context Learning for Data-Scarce Novel Tasks*
Anwoy Chatterjee, Eshaan Tanwar, Subhabrata Dutta and Tanmoy Chakraborty . . . . . 11568

*Split and Rephrase with Large Language Models*
Antonio David Ponce Martínez, Thierry Etchegoyhen, Jesus Javier Calleja Perez and Harritxu Gete . . . . . 11588

*ChunkAttention: Efficient Self-Attention with Prefix-Aware KV Cache and Two-Phase Partition*
Lu Ye, Ze Tao, Yong Huang and Yang Li . . . . . 11608

*AlignBench: Benchmarking Chinese Alignment of Large Language Models*
Xiao Liu, Xuanyu Lei, Shengyuan Wang, Yue Huang, Andrew Zhuoer Feng, Bosi Wen, Jiale Cheng, Pei Ke, Yifan Xu, Weng Lam Tam, Xiaohan Zhang, Lichao Sun, Xiaotao Gu, Hongning Wang, Jing Zhang, Minlie Huang, Yuxiao Dong and Jie Tang . . . . . 11621

*SAPT: A Shared Attention Framework for Parameter-Efficient Continual Learning of Large Language Models*
Weixiang Zhao, Shilong Wang, Yulin Hu, Yanyan Zhao, Bing Qin, Xuanyu Zhang, Qing Yang, Dongliang Xu and Wanxiang Che . . . . . 11641

*DoRA: Enhancing Parameter-Efficient Fine-Tuning with Dynamic Rank Distribution*
Yulong Mao, Kaiyu Huang, Changhao Guan, Ganglin Bao, Fengran Mo and Jinan Xu . . . . 11662

*Cross-Lingual Knowledge Editing in Large Language Models*
Jiaan Wang, Yunlong Liang, Zengkui Sun, Yuxuan Cao, Jiarong Xu and Fandong Meng . . 11676

*Argument Mining in Data Scarce Settings: Cross-lingual Transfer and Few-shot Techniques*
Anar Yeginbergen, Maite Oronoz and Rodrigo Agerri . . . . . 11687

*Learning Task Decomposition to Assist Humans in Competitive Programming*
Jiaxin Wen, Ruiqi Zhong, Pei Ke, Zhihong Shao, Hongning Wang and Minlie Huang . . . . . 11700

*An Entropy-based Text Watermarking Detection Method*
Yijian LU, Aiwei Liu, Dianzhi Yu, Jingjing Li and Irwin King . . . . . 11724

*Enhancing Explainable Rating Prediction through Annotated Macro Concepts*
Huachi Zhou, Shuang Zhou, Hao Chen, Ninghao Liu, Fan Yang and Xiao Huang . . . . . 11736

*How to Engage your Readers? Generating Guiding Questions to Promote Active Reading*
Peng Cui, Vilém Zouhar, Xiaoyu Zhang and Mrinmaya Sachan . . . . . 11749

*Less is More: Mitigating Multimodal Hallucination from an EOS Decision Perspective*
Zihao Yue, Liang Zhang and Qin Jin . . . . . 11766

*Integrate the Essence and Eliminate the Dross: Fine-Grained Self-Consistency for Free-Form Language Generation*
Xinglin Wang, Yiwei Li, Shaoxiong Feng, Peiwen Yuan, Boyuan Pan, Heda Wang, Yao Hu and Kan Li . . . . . 11782

*More frequent verbs are associated with more diverse valency frames: Efficient principles at the lexicon-grammar interface*
Siyu Tao, Lucia Donatelli and Michael Hahn . . . . . 11795

*Quantifying Generalizations: Exploring the Divide Between Human and LLMs' Sensitivity to Quantification*
Claudia Collacciani, Giulia Rambelli and Marianna Bolognesi . . . . . 11811

*Can Large Language Models Interpret Noun-Noun Compounds? A Linguistically-Motivated Study on Lexicalized and Novel Compounds*
Giulia Rambelli, Emmanuele Chersoni, Claudia Collacciani and Marianna Bolognesi . . . . . 11823

*CharacterEval: A Chinese Benchmark for Role-Playing Conversational Agent Evaluation*
Quan Tu, Shilong Fan, Zihang Tian, Tianhao Shen, Shuo Shang, Xin Gao and Rui Yan . . . 11836

*Generative Cross-Modal Retrieval: Memorizing Images in Multimodal Language Models for Retrieval and Beyond*
Yongqi Li, Wenjie Wang, Leigang Qu, Liqiang Nie, Wenjie Li and Tat-Seng Chua . . . . . 11851

*Self-Training with Pseudo-Label Scorer for Aspect Sentiment Quad Prediction*
Yice Zhang, Jie Zeng, Weiming Hu, Ziyi Wang, Shiwei Chen and Ruifeng Xu . . . . . 11862

*Learning to Generate Answers with Citations via Factual Consistency Models*
Rami Aly, Zhiqiang Tang, Samson Tan and George Karypis . . . . . 11876

*Improving Text Embeddings with Large Language Models*
Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder and Furu Wei . 11897

*Self-Training with Direct Preference Optimization Improves Chain-of-Thought Reasoning*
Tianduo Wang, Shichen Li and Wei Lu . . . . . 11917

*UltraLink: An Open-Source Knowledge-Enhanced Multilingual Supervised Fine-tuning Dataset*
Haoyu Wang, Shuo Wang, Yukun Yan, Xujia Wang, Zhiyu Yang, Yuzhuang Xu, Zhenghao Liu, Liner Yang, Ning Ding, Xu Han, Zhiyuan Liu and Maosong Sun . . . . . 11929

*Document-level Claim Extraction and Decontextualisation for Fact-Checking*
Zhenyun Deng, Michael Sejr Schlichtkrull and Andreas Vlachos . . . . . 11943

*PairCFR: Enhancing Model Training on Paired Counterfactually Augmented Data through Contrastive Learning*
Xiaoqi Qiu, Yongjie Wang, Xu Guo, Zhiwei Zeng, Yu Yue, Yuhong Feng and Chunyan Miao 11955

*LLMs Learn Task Heuristics from Demonstrations: A Heuristic-Driven Prompting Strategy for Document-Level Event Argument Extraction*
Hanzhang Zhou, Junlang Qian, Zijian Feng, LU Hui, Zixiao Zhu and Kezhi Mao . . . . . 11972

*Investigating and Mitigating the Multimodal Hallucination Snowballing in Large Vision-Language Models*
Weihong Zhong, Xiaocheng Feng, Liang Zhao, Qiming Li, Lei Huang, Yuxuan Gu, Weitao Ma, Yuan Xu and Bing Qin . . . . . 11991

*mCoT: Multilingual Instruction Tuning for Reasoning Consistency in Language Models*
Huiyuan Lai and Malvina Nissim . . . . . 12012

*GunStance: Stance Detection for Gun Control and Gun Regulation*
Nikesh Gyawali, Iustin Sirbu, Tiberiu Sosea, Sarthak Khanal, Doina Caragea, Traian Rebedea and Cornelia Caragea . . . . . 12027

*Beyond Traditional Benchmarks: Analyzing Behaviors of Open LLMs on Data-to-Text Generation*
Zdeněk Kasner and Ondrej Dusek . . . . . 12045

*Don't Go To Extremes: Revealing the Excessive Sensitivity and Calibration Limitations of LLMs in Implicit Hate Speech Detection*
Min Zhang, Jianfeng He, Taoran Ji and Chang-Tien Lu . . . . . 12073

*Don't Rank, Combine! Combining Machine Translation Hypotheses Using Quality Estimation*
Giorgos Vernikos and Andrei Popescu-Belis . . . . . 12087

*Generating and Evaluating Plausible Explanations for Knowledge Graph Completion*
Antonio Di Mauro, Zhao Xu, Wiem Ben Rim, Timo Sztyler and Carolin Lawrence . . . . . 12106

*One Prompt To Rule Them All: LLMs for Opinion Summary Evaluation*
Tejpalsingh Siledar, Swaroop Nath, Sankara Sri Raghava Ravindra Muddu, Rupasai Rangaraju, Swaprava Nath, Pushpak Bhattacharyya, Suman Banerjee, Amey Patil, Sudhanshu Shekhar Singh, Muthusamy Chelliah and Nikesh Garera . . . . . 12119

*LANDeRMT: Dectecting and Routing Language-Aware Neurons for Selectively Finetuning LLMs to Machine Translation*
Shaolin Zhu, Leiyu Pan, Bo Li and Deyi Xiong . . . . . 12135

*A Joint Coreference-Aware Approach to Document-Level Target Sentiment Analysis*
Hongjie Cai, Heqing Ma, Jianfei Yu and Rui Xia . . . . . 12149

*VisDiaHalBench: A Visual Dialogue Benchmark For Diagnosing Hallucination in Large Vision-Language Models*
Qingxing Cao, Junhao Cheng, Xiaodan Liang and Liang Lin . . . . . 12161

*AutoDSL: Automated domain-specific language design for structural representation of procedures with constraints*
Yu-Zhe Shi, Haofei Hou, Zhangqian Bi, Fanxu Meng, Xiang Wei, Lecheng Ruan and Qining Wang . . . . . 12177

*Multipath parsing in the brain*
Berta Franzluebbers, Donald Dunagan, Miloš Stanojević, Jan Buys and John T. Hale . . . . . 12215

*Search-Adaptor: Embedding Customization for Information Retrieval*
Jinsung Yoon, Yanfei Chen, Sercan O Arik and Tomas Pfister . . . . . 12230

*Back to Basics: Revisiting REINFORCE-Style Optimization for Learning from Human Feedback in LLMs*
Arash Ahmadian, Chris Cremer, Matthias Gallé, Marzieh Fadaee, Julia Kreutzer, Olivier Pietquin, Ahmet Üstün and Sara Hooker . . . . . 12248

*VIEScore: Towards Explainable Metrics for Conditional Image Synthesis Evaluation*
Max Ku, Dongfu Jiang, Cong Wei, Xiang Yue and Wenhu Chen . . . . . 12268

*Tree Transformer's Disambiguation Ability of Prepositional Phrase Attachment and Garden Path Effects*
Lingling Zhou, Suzan Verberne and Gijs Wijnholds . . . . . 12291

*Tree-of-Traversals: A Zero-Shot Reasoning Algorithm for Augmenting Black-box Language Models with Knowledge Graphs*
Elan Sopher Markowitz, Anil Ramakrishna, Jwala Dhamala, Ninareh Mehrabi, Charith Peris, Rahul Gupta, Kai-Wei Chang and Aram Galstyan . . . . . 12302

*Structured Tree Alignment for Evaluation of (Speech) Constituency Parsing*
Freda Shi, Kevin Gimpel and Karen Livescu . . . . . 12320

*ViSAGe: A Global-Scale Analysis of Visual Stereotypes in Text-to-Image Generation*
Akshita Jha, Vinodkumar Prabhakaran, Remi Denton, Sarah Laszlo, Shachi Dave, Rida Qadri, Chandan K. Reddy and Sunipa Dev . . . . . 12333

*Transferable and Efficient Non-Factual Content Detection via Probe Training with Offline Consistency Checking*
Xiaokang Zhang, Zijun Yao, Jing Zhang, Kaifeng Yun, Jifan Yu, Juanzi Li and Jie Tang . . . 12348

*What Do Language Models Learn in Context? The Structured Task Hypothesis.*
Jiaoda Li, Yifan Hou, Mrinmaya Sachan and Ryan Cotterell . . . . . 12365

*Agent Lumos: Unified and Modular Training for Open-Source Language Agents*
Da Yin, Faeze Brahman, Abhilasha Ravichander, Khyathi Chandu, Kai-Wei Chang, Yejin Choi and Bill Yuchen Lin . . . . . 12380

*Investigating Cultural Alignment of Large Language Models*
Badr AlKhamissi, Muhammad ElNokrashy, Mai Alkhamissi and Mona T. Diab . . . . . 12404

*More Victories, Less Cooperation: Assessing Cicero's Diplomacy Play*
Wichayaporn Wongkamjan, Feng Gu, Yanze Wang, Ulf Hermjakob, Jonathan May, Brandon M. Stewart, Jonathan K. Kummerfeld, Denis Peskoff and Jordan Lee Boyd-Graber . . . . . 12423

*VoiceCraft: Zero-Shot Speech Editing and Text-to-Speech in the Wild*
Puyuan Peng, Po-Yao Huang, Shang-Wen Li, Abdelrahman Mohamed and David Harwath 12442

*RAID: A Shared Benchmark for Robust Evaluation of Machine-Generated Text Detectors*
Liam Dugan, Alyssa Hwang, Filip Trhlík, Andrew Zhu, Josh Magnus Ludan, Hainiu Xu, Daphne Ippolito and Chris Callison-Burch . . . . . 12463

*Silent Signals, Loud Impact: LLMs for Word-Sense Disambiguation of Coded Dog Whistles*
Julia Kruk, Michela Marchini, Rijul Magu, Caleb Ziems, David Muchlinski and Diyi Yang12493

*On the Representational Capacity of Neural Language Models with Chain-of-Thought Reasoning*
Franz Nowak, Anej Svete, Alexandra Butoi and Ryan Cotterell . . . . . 12510

*Analyzing LLM Behavior in Dialogue Summarization: Unveiling Circumstantial Hallucination Trends*
Sanjana Ramprasad, Elisa Ferracane and Zachary Chase Lipton . . . . . 12549

*LLM in a flash: Efficient Large Language Model Inference with Limited Memory*
Keivan Alizadeh, Seyed Iman Mirzadeh, Dmitry Belenko, S. Karen Khatamifard, Minsik Cho, Carlo C Del Mundo, Mohammad Rastegari and Mehrdad Farajtabar . . . . . 12562

*Video-ChatGPT: Towards Detailed Video Understanding via Large Vision and Language Models*
Muhammad Maaz, Hanoona Abdul Rasheed, Salman Khan and Fahad Shahbaz Khan .... 12585

*To Distill or Not to Distill? On the Robustness of Robust Knowledge Distillation*
Abdul Waheed, Karima Kadaoui and Muhammad Abdul-Mageed ..... 12603

*LayerSkip: Enabling Early Exit Inference and Self-Speculative Decoding*
Mostafa Elhoushi, Akshat Shrivastava, Diana Liskovich, Basil Hosmer, Bram Wasti, Liangzhen Lai, Anas Mahmoud, Bilge Acun, Saurabh Agarwal, Ahmed Roman, Ahmed A Aly, Beidi Chen and Carole-Jean Wu ..... 12622

*Classist Tools: Social Class Correlates with Performance in NLP*
Amanda Cercas Curry, Giuseppe Attanasio, Zeerak Talat and Dirk Hovy ..... 12643

*ActionIE: Action Extraction from Scientific Literature with Programming Languages*
Xianrui Zhong, Yufeng Du, Siru Ouyang, Ming Zhong, Tingfeng Luo, Qirong Ho, Hao Peng, Heng Ji and Jiawei Han ..... 12656

*A Community-Centric Perspective for Characterizing and Detecting Anti-Asian Violence-Provoking Speech*
Gaurav Verma, Rynaa Grover, Jiawei Zhou, Binny Mathew, Jordan Kraemer, Munmun De Choudhury and Srijan Kumar ..... 12672

*Retaining Key Information under High Compression Ratios: Query-Guided Compressor for LLMs*
Zhiwei Cao, Qian Cao, Yu Lu, Ningxin Peng, Luyang Huang, Shanbo Cheng and Jinsong Su 12685

*COSMIC: Mutual Information for Task-Agnostic Summarization Evaluation*
Maxime Darrin, Philippe Formont, Jackie CK Cheung and Pablo Piantanida ..... 12696

*EUROPA: A Legal Multilingual Keyphrase Generation Dataset*
Olivier Salaün, Frédéric Piedboeuf, Guillaume Le Berre, David Alfonso-Hermelo and Philippe Langlais ..... 12718

*GLIMPSE: Pragmatically Informative Multi-Document Summarization for Scholarly Reviews*
Maxime Darrin, Ines Arous, Pablo Piantanida and Jackie CK Cheung ..... 12737

*Peacock: A Family of Arabic Multimodal Large Language Models and Benchmarks*
Fakhraddin Alwajih, El Moatez Billah Nagoudi, Gagan Bhatia, Abdelrahman Mohamed and Muhammad Abdul-Mageed ..... 12753

*Generating Coherent Sequences of Visual Illustrations for Real-World Manual Tasks*
João Bordalo, Vasco Ramos, Rodrigo Valério, Diogo Glória-Silva, Yonatan Bitton, Michal Yarom, Idan Szpektor and Joao Magalhaes ..... 12777

*Cheetah: Natural Language Generation for 517 African Languages*
Ife Adebara, AbdelRahim A. Elmadany and Muhammad Abdul-Mageed ..... 12798

*TaPERA: Enhancing Faithfulness and Interpretability in Long-Form Table QA by Content Planning and Execution-based Reasoning*
Yilun Zhao, Lyuhao Chen, Arman Cohan and Chen Zhao ..... 12824

*KnowledgeFMath: A Knowledge-Intensive Math Reasoning Dataset in Finance Domains*
Yilun Zhao, Hongjun Liu, Yitao Long, Rui Zhang, Chen Zhao and Arman Cohan ..... 12841

*API-BLEND: A Comprehensive Corpora for Training and Benchmarking API LLMs*
Kinjal Basu, Ibrahim Abdelaziz, Subhajit Chaudhury, Soham Dan, Maxwell Crouse, Asim Munawar, Vernon Austel, Sadhana Kumaravel, Vinod Muthusamy, Pavan Kapanipathi and Luis A. Lastras 12859

*LoRA-Flow: Dynamic LoRA Fusion for Large Language Models in Generative Tasks*
Hanqing Wang, Bowen Ping, Shuo Wang, Xu Han, Yun Chen, Zhiyuan Liu and Maosong Sun 12871

*Harder Task Needs More Experts: Dynamic Routing in MoE Models*
Quzhe Huang, Zhenwei An, Nan Zhuang, Mingxu Tao, Chen Zhang, Yang Jin, Kun Xu, Kun Xu, Liwei Chen, Songfang Huang and Yansong Feng . . . . . 12883

*XLAVS-R: Cross-Lingual Audio-Visual Speech Representation Learning for Noise-Robust Speech Perception*
HyoJung Han, Mohamed Anwar, Juan Pino, Wei-Ning Hsu, Marine Carpuat, Bowen Shi and Changhan Wang . . . . . 12896

*SOTOPIA-: Interactive Learning of Socially Intelligent Language Agents*
Ruiyi Wang, Haofei Yu, Wenxin Sharon Zhang, Zhengyang Qi, Maarten Sap, Yonatan Bisk, Graham Neubig and Hao Zhu . . . . . 12912

*$\mathcal{X}$FT: Unlocking the Power of Code Instruction Tuning by Simply Merging Upcycled Mixture-of-Experts*
Yifeng Ding, Jiawei Liu, Yuxiang Wei and Lingming Zhang . . . . . 12941

*Generalizability of Mixture of Domain-Specific Adapters from the Lens of Signed Weight Directions and its Application to Effective Model Pruning*
Tuc Van Nguyen and Thai Le . . . . . 12956

*Learning to Decode Collaboratively with Multiple Language Models*
Zejiang Shen, Hunter Lang, Bailin Wang, Yoon Kim and David Sontag . . . . . 12974

*DRAGIN: Dynamic Retrieval Augmented Generation based on the Real-time Information Needs of Large Language Models*
Weihang Su, Yichen Tang, Qingyao Ai, Zhijing Wu and Yiqun Liu . . . . . 12991

*Living in the Moment: Can Large Language Models Grasp Co-Temporal Reasoning?*
Zhaochen Su, Juntao Li, Jun Zhang, Tong Zhu, Xiaoye Qu, Pan Zhou, Yan Bowen, Yu Cheng and Min Zhang . . . . . 13014

*CritiqueLLM: Towards an Informative Critique Generation Model for Evaluation of Large Language Model Generation*
Pei Ke, Bosi Wen, Andrew Zhuoer Feng, Xiao Liu, Xuanyu Lei, Jiale Cheng, Shengyuan Wang, Aohan Zeng, Yuxiao Dong, Hongning Wang, Jie Tang and Minlie Huang . . . . . 13034

*LLMArena: Assessing Capabilities of Large Language Models in Dynamic Multi-Agent Environments*
Junzhe Chen, Xuming Hu, Shuodi Liu, Shiyu Huang, Wei-Wei Tu, Zhaofeng He and Lijie Wen 13055

*Small But Funny: A Feedback-Driven Approach to Humor Distillation*
Sahithya Ravi, Patrick Huber, Akshat Shrivastava, Vered Shwartz and Arash Einolghozati . 13078

*Symbol-LLM: Towards Foundational Symbol-centric Interface For Large Language Models*
Fangzhi Xu, Zhiyong Wu, Qiushi Sun, Siyu Ren, Fei Yuan, Shuai Yuan, Qika Lin, Yu Qiao and Jun Liu . . . . . 13091

*From Sights to Insights: Towards Summarization of Multimodal Clinical Documents*
Akash Ghosh, Mohit Singh Tomar, Abhisek Tiwari, Sriparna Saha, Jatin Avinash Salve and Setu Sinha . . . . . 13117

*When Phrases Meet Probabilities: Enabling Open Relation Extraction with Cooperating Large Language Models*
Jiaxin Wang, Lingling Zhang, Wee Sun Lee, Yujie Zhong, Liwei Kang and Jun Liu . . . . . 13130

*Effects of diversity incentives on sample diversity and downstream model performance in LLM-based text augmentation*
Jan Cegin, Branislav Pecher, Jakub Simko, Ivan Srba, Maria Bielikova and Peter Brusilovsky 13148

*Beyond Orthography: Automatic Recovery of Short Vowels and Dialectal Sounds in Arabic*
Yassine El Kheir, Hamdy Mubarak, Ahmed Ali and Shammur Absar Chowdhury . . . . . 13172

*Document-Level Machine Translation with Large-Scale Public Parallel Corpora*
Proyag Pal, Alexandra Birch and Kenneth Heafield . . . . . 13185

*Bridging the Empirical-Theoretical Gap in Neural Network Formal Language Learning Using Minimum Description Length*
Nur Lan, Emmanuel Chemla and Roni Katzir . . . . . 13198

*Context versus Prior Knowledge in Language Models*
Kevin Du, Vésteinn Snæbjarnarson, Niklas Stoehr, Jennifer C. White, Aaron Schein and Ryan Cotterell . . . . . 13211

*Word Matters: What Influences Domain Adaptation in Summarization?*
Yinghao Li, Siyu Miao, Heyan Huang and Yang Gao . . . . . 13236

*Visualization Recommendation with Prompt-based Reprogramming of Large Language Models*
Xinhang Li, Jingbo Zhou, Wei Chen, Derong Xu, Tong Xu and Enhong Chen . . . . . 13250

*HOLMES: Hyper-Relational Knowledge Graphs for Multi-hop Question Answering using LLMs*
Pranoy Panda, Ankush Agarwal, Chaitanya Devaguptapu, Manohar Kaul and Prathosh AP 13263

*Toward In-Context Teaching: Adapting Examples to Students' Misconceptions*
Alexis Ross and Jacob Andreas . . . . . 13283

*Bridging Word-Pair and Token-Level Metaphor Detection with Explainable Domain Mining*
Yuan Tian, Ruike Zhang, Nan Xu and Wenji Mao . . . . . 13311

*Faithful Logical Reasoning via Symbolic Chain-of-Thought*
Jundong Xu, Hao Fei, Liangming Pan, Qian Liu, Mong-Li Lee and Wynne Hsu . . . . . 13326

*$S^2$GSL: Incorporating Segment to Syntactic Enhanced Graph Structure Learning for Aspect-based Sentiment Analysis*
Bingfeng Chen, Qihan Ouyang, Yongqi Luo, Boyan Xu, Ruichu Cai and Zhifeng Hao . . . . 13366

*Maverick: Efficient and Accurate Coreference Resolution Defying Recent Trends*
Giuliano Martinelli, Edoardo Barba and Roberto Navigli . . . . . 13380

*ESCoT: Towards Interpretable Emotional Support Dialogue Systems*
Tenggan Zhang, Xinjie Zhang, Jinming Zhao, Li Zhou and Qin Jin . . . . . 13395

*PathReasoner: Modeling Reasoning Path with Equivalent Extension for Logical Question Answering*
Fangzhi Xu, Qika Lin, Tianzhe Zhao, JiaweiHan JiaweiHan and Jun Liu . . . . . 13413

*WARDEN: Multi-Directional Backdoor Watermarks for Embedding-as-a-Service Copyright Protection*
Anudeex Shetty, Yue Teng, Ke He and Qiongkai Xu . . . . . 13430

*Advancing Parameter Efficiency in Fine-tuning via Representation Editing*
Muling Wu, Wenhao Liu, Xiaohua Wang, Tianlong Li, Changze Lv, Zixuan Ling, Zhu JianHao, Cenyuan Zhang, Xiaoqing Zheng and Xuanjing Huang . . . . . 13445

*Context Consistency between Training and Inference in Simultaneous Machine Translation*
Meizhi Zhong, Lemao Liu, Kehai Chen, Mingming Yang and Min Zhang . . . . . 13465

*Using Natural Language Explanations to Improve Robustness of In-context Learning*
Xuanli He, Yuxiang Wu, Oana-Maria Camburu, Pasquale Minervini and Pontus Stenetorp 13477

*Chunk, Align, Select: A Simple Long-sequence Processing Method for Transformers*
Jiawen Xie, Pengyu Cheng, Xiao Liang, Yong Dai and Nan du . . . . . 13500

*ArchCode: Incorporating Software Requirements in Code Generation with Large Language Models*
Hojae Han, Jaejin Kim, Jaeseok Yoo, Youngwon Lee and Seung-won Hwang . . . . . 13520

*Combining Supervised Learning and Reinforcement Learning for Multi-Label Classification Tasks with Partial Labels*
Zixia Jia, Junpeng Li, Shichuan Zhang, Anji Liu and Zilong Zheng . . . . . 13553

*MULFE: A Multi-Level Benchmark for Free Text Model Editing*
Chenhao Wang, Pengfei Cao, Zhuoran Jin, Yubo Chen, Daojian Zeng, Kang Liu and Jun Zhao 13570

*MobileSpeech: A Fast and High-Fidelity Framework for Mobile Zero-Shot Text-to-Speech*
Shengpeng Ji, Ziyue Jiang, Hanting Wang, Jialong Zuo and Zhou Zhao . . . . . 13588

*Spatially-Aware Speaker for Vision-and-Language Navigation Instruction Generation*
Muraleekrishna Gopinathan, Martin Masek, Jumana Abu-Khalaf and David Suter . . . . . 13601

*HiRoPE: Length Extrapolation for Code Models Using Hierarchical Position*
Kechi Zhang, Ge Li, Huangzhao Zhang and Zhi Jin . . . . . 13615

*Never Lost in the Middle: Mastering Long-Context Question Answering with Position-Agnostic Decompositional Training*
Junqing He, Kunhao Pan, Xiaoqun Dong, Zhuoyang Song, LiuYiBo LiuYiBo, Qianguosun Qianguosun, Yuxin Liang, Hao Wang, Enming Zhang and Jiaxing Zhang . . . . . 13628

*CodeAgent: Enhancing Code Generation with Tool-Integrated Agent Systems for Real-World Repo-level Coding Challenges*
Kechi Zhang, Jia Li, Ge Li, Xianjie Shi and Zhi Jin . . . . . 13643

*When is Tree Search Useful for LLM Planning? It Depends on the Discriminator*
Ziru Chen, Michael White, Ray Mooney, Ali Payani, Yu Su and Huan Sun . . . . . 13659

*LogicBench: Towards Systematic Evaluation of Logical Reasoning Ability of Large Language Models*
Mihir Parmar, Nisarg Patel, Neeraj Varshney, Mutsumi Nakamura, Man Luo, Santosh Mashetty, Arindam Mitra and Chitta Baral . . . . . 13679

*Meta-Tuning LLMs to Leverage Lexical Knowledge for Generalizable Language Style Understanding*
Ruohao Guo, Wei Xu and Alan Ritter . . . . . 13708

*Reducing Privacy Risks in Online Self-Disclosures with Language Models*
Yao Dou, Isadora Krsek, Tarek Naous, Anubha Kabra, Sauvik Das, Alan Ritter and Wei Xu13732

*Navigating the Dual Facets: A Comprehensive Evaluation of Sequential Memory Editing in Large Language Models*
Zihao Lin, Mohammad Beigi, Hongxuan Li, Yufan Zhou, Yuxiang Zhang, Qifan Wang, Wenpeng Yin and Lifu Huang . . . . . 13755

*REFINESUMM: Self-Refining MLLM for Generating a Multimodal Summarization Dataset*
Vaidehi Patil, Leonardo F. R. Ribeiro, Mengwen Liu, Mohit Bansal and Markus Dreyer . . . 13773

*When Benchmarks are Targets: Revealing the Sensitivity of Large Language Model Leaderboards*
Norah A. Alzahrani, Hisham Abdullah Alyahya, Yazeed Alnumay, Sultan AlRashed, Shaykhah Z. Alsubaie, Yousef Almushayqih, Faisal Abdulrahman Mirza, Nouf M. Alotaibi, Nora Al-Twairesh, Areeb Alowisheq, M Saiful Bari and Haidar Khan . . . . . 13787

*LLM-Rubric: A Multidimensional, Calibrated Approach to Automated Evaluation of Natural Language Texts*
Helia Hashemi, Jason Eisner, Corby Rosset, Benjamin Van Durme and Chris Kedzie . . . . . 13806

*LIEDER: Linguistically-Informed Evaluation for Discourse Entity Recognition*
Xiaomeng Zhu and Robert Frank . . . . . 13835

*Evaluating Very Long-Term Conversational Memory of LLM Agents*
Adyasha Maharana, Dong-Ho Lee, Sergey Tulyakov, Mohit Bansal, Francesco Barbieri and Yuwei Fang . . . . . 13851

*Prototypical Reward Network for Data-Efficient Model Alignment*
Jinghan Zhang, Xiting Wang, Yiqiao Jin, Changyu Chen, Xinhao Zhang and Kunpeng Liu 13871

*NEO-BENCH: Evaluating Robustness of Large Language Models with Neologisms*
Jonathan Zheng, Alan Ritter and Wei Xu . . . . . 13885

*Impacts of Misspelled Queries on Translation and Product Search*
Greg Hanneman, Natawut Monaikul and Taichi Nakatani . . . . . 13907

*Skin-in-the-Game: Decision Making via Multi-Stakeholder Alignment in LLMs*
Bilgehan Sel, Priya Shanmugasundaram, Mohammad Kachuee, Kun Zhou, Ruoxi Jia and Ming Jin . . . . . 13921

*The MERSA Dataset and a Transformer-Based Approach for Speech Emotion Recognition*
Enshi Zhang, Rafael Trujillo and Christian Poellabauer . . . . . 13960

*Transparent and Scrutable Recommendations Using Natural Language User Profiles*
Jerome Ramos, Hossein A. Rahmani, Xi Wang, Xiao Fu and Aldo Lipani . . . . . 13971

*Fora: A corpus and framework for the study of facilitated dialogue*
Hope Schroeder, Deb Roy and Jad Kabbara . . . . . 13985

*Explanation-aware Soft Ensemble Empowers Large Language Model In-context Learning*
Yue Yu, Jiaming Shen, Tianqi Liu, Zhen Qin, Jing Nathan Yan, Jialu Liu, Chao Zhang and Michael Bendersky . . . . . 14002

*What is the Best Way for ChatGPT to Translate Poetry?*
Shanshan Wang, Derek F. Wong, Jingming Yao and Lidia S. Chao . . . . . 14025

*Rephrasing the Web: A Recipe for Compute and Data-Efficient Language Modeling*
Pratyush Maini, Skyler Seto, Richard He Bai, David Grangier, Yizhe Zhang and Navdeep Jaitly 14044

*DeCoT: Debiasing Chain-of-Thought for Knowledge-Intensive Tasks in Large Language Models via Causal Intervention*
Junda Wu, Tong Yu, Xiang Chen, Haoliang Wang, Ryan A. Rossi, Sungchul Kim, Anup Rao and Julian McAuley . . . . . 14073

*Representation Learning with Conditional Information Flow Maximization*
Dou Hu, Lingwei Wei, Wei Zhou and Songlin Hu . . . . . 14088

*GPT is Not an Annotator: The Necessity of Human Annotation in Fairness Benchmark Construction*
Virginia K. Felkner, Jennifer A. Thompson and Jonathan May . . . . . 14104

*Quantifying Contamination in Evaluating Code Generation Capabilities of Language Models*
Martin Riddell, Ansong Ni and Arman Cohan . . . . . 14116

*Language Models are Homer Simpson! Safety Re-Alignment of Fine-tuned Language Models through Task Arithmetic*
Rishabh Bhardwaj, Duc Anh Do and Soujanya Poria . . . . . 14138

*Tracking the Newsworthiness of Public Documents*
Alexander Spangher, Serdar Tumgoren, Ben Welsh, Nanyun Peng, Emilio Ferrara and Jonathan May . . . . . 14150

*EWEK-QA : Enhanced Web and Efficient Knowledge Graph Retrieval for Citation-based Question Answering Systems*
Mohammad Dehghan, Mohammad Ali Alomrani, Sunyam Bagga, David Alfonso-Hermelo, Khalil Bibi, Abbas Ghaddar, Yingxue Zhang, Xiaoguang Li, Jianye Hao, Qun Liu, Jimmy Lin, Boxing Chen, Prasanna Parthasarathi, Mahdi Biparva and Mehdi Rezagholizadeh . . . . . 14169

*Multi-modal Preference Alignment Remedies Degradation of Visual Instruction Tuning on Language Models*
Shengzhi LI, Rongyu Lin and Shichao Pei . . . . . 14188

*Multistage Collaborative Knowledge Distillation from a Large Language Model for Semi-Supervised Sequence Generation*
Jiachen Zhao, Wenlong Zhao, Andrew Drozdov, Benjamin Rozonoyer, Md Arafat Sultan, Jay-Yoon Lee, Mohit Iyyer and Andrew McCallum . . . . . 14201

*Controlled Text Generation for Black-box Language Models via Score-based Progressive Editor*
Sangwon Yu, Changmin Lee, Hojin Lee and Sungroh Yoon . . . . . 14215

*LogogramNLP: Comparing Visual and Textual Representations of Ancient Logographic Writing Systems for NLP*
Danlu Chen, Freda Shi, Aditi Agarwal, Jacobo Myerston and Taylor Berg-Kirkpatrick . . . . 14238

*Superfiltering: Weak-to-Strong Data Filtering for Fast Instruction-Tuning*
Ming Li, Yong Zhang, Shwai He, Zhitao Li, Hongyu Zhao, Jianzong Wang, Ning Cheng and Tianyi Zhou . . . . . 14255

*Confabulation: The Surprising Value of Large Language Model Hallucinations*
Peiqi Sui, Eamon Duede, Sophie Wu and Richard Jean So . . . . . 14274

*IAPT: Instance-Aware Prompt Tuning for Large Language Models*
Wei Zhu, Aaron Xuxiang Tian, Congrui Yin, Yuan Ni, Xiaoling Wang and Guotong Xie . . 14285

# Program

**Saturday, August 10, 2024**

14:00 - 19:00 *Registration*

**Sunday, August 11, 2024**

07:30 - 20:30 *Registration Check in*

10:30 - 11:00 *Break*

12:30 - 14:00 *Box Lunch Provided*

15:30 - 16:00 *Break*

18:30 - 21:00 *Welcome Reception*

**Monday, August 12, 2024**

07:30 - 16:30 *Registration*

09:00 - 10:30 *Session 1: Plenary - Opening Session & Keynote TBA*

10:30 - 11:00 *Break*

11:00 - 12:30 *Session 2: Oral's/Posters/ Demo Presentations A*

12:45 - 13:45 *Session 3: Finding Presentations 1*

14:00 - 15:30 *Session 4: Oral's/Posters/ Demo Presentations B*

15:30 - 16:00 *Break*

16:00 - 17:30 *Session 5: Oral's/Posters/ Demo Presentations C*

17:45 - 18:45 *Session 6: Findings Presentations 2*

**Tuesday, August 13, 2024**

08:30 - 16:30 *Registration*

09:00 - 10:00 *Session 7: Plenary - Keynote TBA*

10:00 - 10:30 *Break*

10:30 - 12:00 *Session 8: Oral's/Posters/ Demo Presentations D*

12:15 - 13:15 *Session 9: Finding Presentations 3*

13:00 - 14:00 *Session 10: Plenary - Business Meeting (all attendees welcome)*

14:30 - 15:30 *Session 11: Plenary - Panel*

15:30 - 16:00 *Break*

16:00 - 17:30 *Session 12: Oral's/Posters/ Demo Presentations E*

19:00 - 22:00 *Social Event Dinner*

**Wednesday, August 14, 2024**

08:30 - 16:30 *Registration*

09:00 - 10:00 *Session 13: Plenary - Keynote TBA*

10:00 - 10:30 *Break*

10:30 - 12:00 *Session 14: Oral's/Posters/ Demo Presentations F*

12:15 - 13:15 *Session 15: Finding Presentations 4*

13:30 - 15:00 *Session 16: Plenary - Lifetime Achievement & ToT Awards*

15:00 - 15:30 *Break*

15:30 - 17:00 *Session 17: Plenary - Paper Awards*

17:15 - 18:00 *Session 18: Plenary - Closing Session*