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

**Diamond**

- Google Research
- Apple
- GTCOM
- Colossal-AI

**Platinum**

- Ahrefs
- Alibaba Cloud
- Amazon | Science
- Baidu

- ByteDance
- Cohere
- Megagon Labs
- NEC

**Gold**

- Ant Group
- Bloomberg Engineering
- Huawei
- JPMorgan Chase & Co.
Silver

aixplain

duolingo

Bronze

Adobe

Abelscape

mercari

nyonic

Diversity & Inclusion Champion

Diversity & Inclusion Ally

amazon | science
©2023 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 979-8-89176-060-8
Message from the General Chair

I am happy to welcome you to EMNLP-2023 in Singapore! Like EMNLP-2021, EMNLP-2022, and other ACL-related meetings, we decided to host EMNLP-2023 as another hybrid conference having both in-person and virtual presentations and participants. We are not sure how long this style of our meetings will last. However, we have already accustomed to this style of conferences, which has its own advantages, while it causes a heavy burden to those organizing such events.

The past one year has been a terrific and thrilling year since the advent of ChatGPT and other Large Language Models. Any people having access to those models has posed a big impact on people’s impression about AI and has started to give them a feeling of fear. People now can do not only natural conversation with AI but also conduct various natural language tasks using our own languages. We now know it is difficult to guarantee that Large Language Models produce honest and harmless outputs. We have found that good prompting, demonstrations and complex ones like the Chain of Thought prompting draw out or enhance the emergent abilities of Large Language Models. However, we still don’t know precisely why and how such in-context learning works. This year’s EMNLP highlights a theme track, “Large Language Models and the Future of NLP.” I hope we can see enthusiastic discussions and innovative ideas will be presented in EMNLP-2023.

One big trial is that the Program Chairs decided to use OpenReview as the cradle of the main conference papers, for making reviews and author responses publicly available. The motivation and effects of this trial will be explained by the PC Chairs.

Another important trial is to rent out the Universal Studio Singapore for our Social Event. I hope everyone will enjoy this event.

EMNLP-2023 is the biggest conference ever in the SIGDAT history. Organizing such a big event is very difficult. As the General Chair, the most important and difficult task is to organize all the committees by a group of enthusiastic and talented people. I was very fortunate to be able to collect great committee members. Without such a wonderful group of colleagues, it almost has been impossible to make this great event happen. I would like to send my sincere thanks to all the members of our organization teams. Here, I only list the chairs by names, but I also like to send gratitude from my heart to all the people involved in EMNLP-2023, including keynote speakers, panelists, workshop organizers, tutorial tutors, senior area chairs, area chairs, reviewers, volunteers, sponsors, the Underline team, and all of you attending EMNLP-2023 in-person or virtually.

• The program chairs — Houda Bouamor, Juan Pino, and Kalika Bali — who made a number of innovations and handled a huge number of submitted papers. I cannot help but be grateful for their tireless work.

• The Local Chair and the Local Team — Haizhou Li the Chair organized and lead a wonderful group of people. While I cannot name every one of them, weekly meetings with the team members including related Chairs made our communication smooth and worked as a good time-keeper.

For the remaining committee chairs, I only list them by names, as I cannot give all my gratitude only with short messages.

• The Industry Track Chairs — Mingxuan Wang and Imed Zitouni

• The Workshop Chairs — Sujian Li, Alex Marin, and Hao Fei

• The Tutorial Chairs — Qi Zhang and Hassan Sajjad
• The Ethics Chairs — Kemal Oflazer and Francisco Guzman
• The Demonstration Chairs — Yansong Feng and Els Lefever
• The Publication Chairs — Nadi Tomeh, Atsushi Fujita, Aixin Sun, Bin Wang, Rong Tong, and Ryan Cotterell
• Publicity Chairs: Weiwei Sun and Noriki Nishida
• The Student Volunteer Chairs — Mamoru Komachi, Kevin Duh, and Ekaterina Kochmar
• The Diversity/Inclusion Chairs — Jing Li, Luisa Bentivogli, Eva Vanmassenhove, and Shi Zong
• The Reviewer Mentoring Chairs — Roi Reichart and Roy Lee
• The Sponsorship Chairs — Deyi Xiong, Shyam Upadhyay, and the ACL Sponsorship Director, Chris Callison-Burch
• The Website Chairs — Yan Zhang and Benyou Wang
• The Virtual Infrastructure Chairs — Feng Jiang, Chen Zhang, and the Underline Science team, especially Damira Mršić

Last but least, I want to express special thanks to Jennifer Rachford, the ACL Business Manager for her endless attention and guidance to all aspects of the organization.

I also express my gratitude to our sponsors, without their donation this conference would not be possible:

• Diamond Sponsors: Apple, Colossal-AI, Google Research, GTCOM, King Salman Global Academy for Arabic Language, LivePerson, and SONY
• Platinum Sponsors: Ahrefs, Alibaba Cloud, Amazon, Baidu, ByteDance, Cohere, Megagon Labs, and NEC
• Gold Sponsors: ANT GROUP, Bloomberg, HUAWEI, J.P.Morgan Chase, Salesforce, and SAP
• Silver Sponsors: aiXplain, duolingo, jenni, and Translated
• Bronze Sponsors: Adobe, Babelscape, ModelBest, and nyonic
• Diversity and Inclusion, Champion: RIKEN AIP
• Diversity and Inclusion, Ally: Amazon

Finally, I hope all of you enjoy the main conference, workshops, tutorials and all other events, and make your stay in Singapore an unforgettable experience!

Yuji Matsumoto
RIKEN AIP, Tokyo, Japan
EMNLP 2023 General Chair
Message from the Program Co-Chairs

Welcome to EMNLP 2023, one of the most-attended conferences in the field of Natural Language Processing, held in “hybrid” mode this year serving both virtual and in-person participants in Singapore, where we have increased the number of Findings papers to be more inclusive and to showcase the diversity and quality of research in our field.

Submission and Acceptance

EMNLP 2023 received 4,909 full paper submissions, the largest number to date. This number includes 134 ARR papers that were committed to EMNLP (see further discussion of ARR below). 256 papers were desk rejected for various reasons (missing limitation section, anonymity policy, multiple submission policy or formatting violations), leaving us with submissions that were fully reviewed. This is a record-breaking number of submissions, an increase of 719 submissions over last year. Based on the reviewers, area chairs and senior area chair comments, we have tried to keep the EMNLP 2023 main acceptance rates similar to previous events while increasing the number of Findings.

We accepted 1,105 papers to the main conference including 9 nominated by the Computational Linguistics (CL) journal, and 49 by the Transactions of the Association of Computational Linguistics (TACL). Out of these, 234 are oral presentations and 871 poster presentations. We also accepted 1,060 papers for “Findings of EMNLP”. Additionally, a total of 77 and 52 submissions were accepted to be presented in the Industry and Demo tracks, respectively.

More statistics on the accepted papers can be found below.

<table>
<thead>
<tr>
<th></th>
<th>Long</th>
<th>Short</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submitted (Including ARR commits)</td>
<td>3,868</td>
<td>1,041</td>
<td>4,909</td>
</tr>
<tr>
<td>Accepted to the Main conference</td>
<td>901</td>
<td>146</td>
<td>1,047</td>
</tr>
<tr>
<td>Acceptance Rate (main conference)</td>
<td>23.3%</td>
<td>14.0%</td>
<td>21.3%</td>
</tr>
<tr>
<td>Accepted to Findings</td>
<td>857</td>
<td>203</td>
<td>1,060</td>
</tr>
<tr>
<td>Acceptance Rate (Main + Findings)</td>
<td>45.4%</td>
<td>33.5%</td>
<td>42.9%</td>
</tr>
<tr>
<td>Presented TACL papers</td>
<td></td>
<td></td>
<td>49</td>
</tr>
<tr>
<td>Presented CL papers</td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Accepted Industry</td>
<td></td>
<td></td>
<td>77</td>
</tr>
<tr>
<td>Accepted Demo</td>
<td></td>
<td></td>
<td>52</td>
</tr>
</tbody>
</table>

Limitations Statement

We continued the practice started in the previous year with a requirement that each submitted paper must include an explicitly named Limitations section, discussing the limitations of the work. This discussion does not count towards the page limit, and was strictly implemented. A large number of desk rejects were due to a missing Limitations section.

Tracks

To ensure a smooth process, the submissions to EMNLP 2023 were divided into 27 tracks. The tracks mostly followed those of previous EMNLP conferences, reflecting the “standard” divisions in the field. We did however make the following changes: the “Multilinguality” track became “Multilinguality and Linguistic Diversity” to explicitly call out the work on diverse languages; and we added ONE additional track on “Human-centered NLP”. Finally, we continued the inclusion of a “Theme Track” and solicited
papers on “Large Language Models and the Future NLP”, discussing empirical and theoretical research, as well as position and survey papers on the ways in which the new generation of Large Language Models perform for NLP tasks and applications, and what this means for the future of NLP as a field. Of the 27 tracks, the Resources and Evaluation, NLP Applications, Dialogue and Interactive Systems, Information Extraction, and the Theme tracks were the most popular with over 200 submissions per track.

Program committee structure & reviewing

Similar to prior NLP conferences, we adopted the hierarchical program committee structure, where for each area we invited between 1 to 5 Senior Area Chairs (SACs), who worked with a team of Area Chairs (ACs), and an army of reviewers. We relied on statistics from prior years to estimate how many SACs, ACs and reviewers will be needed and ended up with 85 SACs and 458 ACs. For the reviewers, we used the reviewer lists from prior conferences, solicited volunteer reviewers. To this end, we used the reviewers from ACL2023. We provided the information to the program committee for making reviewer assignments. This resulted in a reviewer pool of 4,094 reviewers of which at least 3,643 reviewers submitted at least one review. For each submission, we assigned three reviewers and an AC. The initial paper assignment was made using an automatic algorithm that matches the abstracts and submitted PDFs with ACs/reviewers’ past publication records, then the assignments were further refined by the SACs/PCs. This matching was done at the global level and not localized per track, which meant that a reviewer could review two (or more) relevant papers from different tracks.

Open Review (see Transparency and Quality of Reviews below) calculated affinity scores making sure that each reviewer does not have a load exceeding 5 papers. We also accommodated requests for a limited load. Besides the overall recommendation, reviewers were asked whether there was any ethical concern. To ensure the review quality, we provided detailed guidelines about what reviewers should and shouldn’t do in a review.

We made final decisions according to the rankings and SAC recommendations. Our final decisions were made not just on the review scores, but also took into account the reviews, author responses, discussion, meta-reviews and SAC/AC recommendations.

Transparency and Quality of Reviews

For a transparent and open reviewing system, this year we have implemented a process by which some of the reviews, author responses and meta reviews will be made publicly available. Our motivation here was to provide increased transparency in the review process, to foster more accountability for reviewers and higher quality reviews as well as enabling peer review research by providing an open collection of papers and reviews. Only reviews, author responses and meta reviews of accepted papers and opt-in rejected papers (where opting in is done by authors) will be made publicly available after the acceptance notifications. The requirement of making reviews open meant that we had to move the conference platform from SoftConf to OpenReview. The OpenReview platform is also currently being used by the ACL Rolling Review and some other related conferences and is well suited to this type of process.

We also introduced a rebuttal and discussions cycle to the reviewing process, where we allowed the reviewers and the authors to have conversations as a part of it. We also allowed for some additional experimentations during the rebuttal. This has, on the whole, seen a positive engagement from the authors and the reviewers. These discussions were extremely useful for the meta-reviews, and final paper decisions. This, we believe, also helps in ensuring high quality reviews to a certain extent.

A lot of behind-the-scenes technical and other work has gone into making this possible. We are grateful to everyone for their patience with some of the glitches on the way. We hope that more and more authors
will choose to make their reviews and data available for the community, and more and more reviewers will engage in the rebuttal discussions.

**Ethics committee**

We also formed an Ethics Committee (EC) dedicated to ethical issues. Besides, we strongly encouraged the authors to include an ethics statement that did not count towards the page limit. After the technical reviews, but before the author-response and discussion phases, the ethics committee considered 54 papers that were flagged by the technical reviewing committee for ethical concerns. The two EC chairs went over the papers, to determine whether a full EC review would be required. As a result, 23 papers received two dedicated ethics reviews from a committee of 42 reviewers recruited by the EC chairs. We thank the committee for their excellent work.

**ACL Rolling Review**

ACL Rolling Review (ARR) is an initiative of the Association for Computational Linguistics, where the reviewing and acceptance of papers to publication venues are done in a two-step process: (1) centralized rolling review and (2) the ability to commit the reviewed papers to be considered for publication by a publication venue. For EMNLP 2023, we continued the process from last year where the authors could either submit papers to EMNLP 2023 directly, or commit ARR reviewed papers by a certain date. We coordinated with the ARR team to extract the submission, review and meta-review. ARR submissions already have their reviews and meta-recommendation. These ARR papers were then ranked by the SACs of the given tracks.

Overall, EMNLP had 134 papers committed from ARR, of these 52 were accepted to the main conference and 45 were accepted to Findings of EMNLP.

**Best paper selection**

This year we increased the number of best paper awards as following:

a) Best Paper (Long)

b) Best Paper (Short)

c) Best Paper (Theme)

d) Best Paper (Industry)

e) Best Paper (Demo)

f) Outstanding Paper in each of the tracks including the Theme, Industry and Demo tracks

Based on the nominations from SACs and ACs, we have identified 107 candidates for consideration for the best papers and outstanding papers award. These papers are assessed by the Best Paper Award Committee. The award winners will be announced and present their works at the closing ceremony.

**Presentation Mode**

We attempted the decision for oral vs poster presentations not to be made based on the quality/merit of the papers, but rather on the authors’ interest in the presentation mode, and our understanding of what would be the best format for presentation of each individual paper.
Keynote talks

Another highlight of our program is the three exciting keynote talks, presented by Prof. Jong Park, KAIST, on “Human-Centric Natural Language Processing”; Prof. Emily M. Provost, University of Michigan, on “From Speech to Emotion to Mood: Mental Health Modeling in Real-World Environments” & Prof. Christopher Manning from Stanford University, on “Academic NLP research in the Age of LLMs: Nothing but blue skies!”

Gratitude

We would like to thank the following people for their support & contributions:

• Our General Chair, Yuji Matsumoto, who led the whole organizing team, and helped with many of the decision processes;

• 85 SACs who helped us throughout the entire review process, from assigning papers, checking review quality, making final recommendation, suggesting presentation formats to recommending best paper candidates;

• 458 ACs who checked the initial submissions, led paper discussions, wrote meta reviews, ensured review quality and suggested best paper candidates;

• 3,643 reviewers who reviewed the papers and actively participated in paper discussions; special thanks to those who stepped in at the last minute to serve as emergency reviewers;

• 42 Ethics Committee members, chaired by Kemal Oflazer and Francisco Guzman, for their hard work in providing ethical reviews and meta-reviews for all papers with serious ethical issues, and ensure that all the conditionally accepted papers have addressed the ethical issues appropriately;

• The 26 members of the Best Paper Award Committee for selecting the best papers;

• Publication Chairs Nadi Tomeh, Atsushi Fujita, Aixin Sun, Bin Wang, Rong Tong, and Ryan Cotterell for completing the final proceedings within a short period;

• ACL Anthology Director Matt Post and his team, for his help in the production of the conference proceedings and maintenance of the ACL Anthology;

• TACL editor-in-chief Asli Celikyilmaz & Roi Reichart, TACL Editorial Assistant Cindy Robinson, and CL Editor-in-Chief Hwee Tou Ng for coordinating TACL and CL presentations with us;

• The ARR team for their continued effort in running ARR, and for coordination with us;

• The Open Review team, especially Harold Rubio and Celeste Matinez for multiple rounds of technical help in setting up EMNLP 2023 on the OR platform;

• Website chairs Yan Zhang and Benyou Wang for their continued effort in prompt updates to the website;

• Publicity chairs Weiwei Sun and Noriki Nishada for publicizing the conference and handling communications on social media.

• Damira Mršic and the whole Underline team, for helping to manage the logistics of both the virtual and online conference.

• Jenn Rachford for her patience, and professional invaluable help in organizing the logistics of the conference.
• Haizhou Li and the rest of the Local Organizing Committee, for various discussions on organizing EMNLP, and making the local arrangements.

• 14,105 authors who submitted their work to EMNLP 2023.

We hope that you will enjoy this year’s program and hybrid conference!

Houda Bouamor, Carnegie Mellon University in Qatar
Juan Pino, Meta
Kalika Bali, Microsoft Research Labs India
EMNLP 2023 Program Co-Chairs
Organizing Committee

General Chair
Yuji Matsumoto, RIKEN Center for Advanced Intelligence Project

Program Chairs
Houda Bouamor, Carnegie Mellon University in Qatar
Juan Pino, Meta
Kalika Bali, Microsoft Research Labs India

Local Chair
Haizhou Li, The Chinese University of Hong Kong, Shenzhen & National University of Singapore

Tutorial Chairs
Qi Zhang, Fudan University
Hassan Sajjad, Dalhousie University

Workshop Chairs
Sujian Li, Peking University
Alex Marin, Microsoft
Hao Fei, National University of Singapore

Demonstration Chairs
Yansong Feng, Peking University
Els Lefever, LT3, Ghent University

Industry Track Chairs
Mingxuan Wang, ByteDance AI Lab
Imed Zitouni, Google

Ethics Chairs
Kemal Oflazer, CMU
Francisco Guzman, Meta

Publication Chairs
Nadi Tomeh, Université Sorbonne Paris Nord
Atsushi Fujita, NICT
Aixin Sun, Nanyang Technological University
Bin Wang, Institute for Infocomm Research, A*STAR, Singapore
Rong Tong, Singapore Institute of Technology

xii
Ryan Cotterell, ETH Zürich

Publicity and Social Media Chairs
Weiwei Sun, University of Cambridge
Noriki Nishida, RIKEN AIP

Website Chairs
Yan Zhang, National University of Singapore
Benyou Wang, The Chinese University of Hong Kong, Shenzhen

Student Volunteer Chairs
Mamoru Komachi, Hitotsubashi University
Kevin Duh, Johns Hopkins University
Ekaterina Kochmar, MBZUAI

Virtual Infrastructure Chairs
Feng Jiang, The Chinese University of Hong Kong, Shenzhen
Chen Zhang, National University of Singapore

Diversity and Inclusion Chairs
Jing Li, The Hong Kong Polytechnic University
Luisa Bentivogli, Fondazione Bruno Kessler
Eva Vanmassenhove, Tilburg University
Shi Zong, University of Waterloo

Reviewer Mentoring Chairs
Roi Reichart, Technion
Roy Lee, Singapore University of Technology and Design

Sponsorship Chairs
Deyi Xiong, Tianjin University
Shyam Upadhyay, Google

ACL Event Director
Jennifer Rachford

ACL On-Site Team
Brandy Dorsey
Megan Maloy
Sally Stevenson
AV Team - Lee Hartman & Sons

Jon Dorsey
Trevor Laffoon

Virtual Hybrid Team - Underline

Nateo Antonic
Borna Bevanda
Dorian Fildor
Rafael Grabovica
Jernej Masnec
Damira Mrsic
Alexandru Pricop
Lucijan Prpic
Luka Simic
Program Committee

Commonsense Reasoning
Niranjan Balasubramanian, Antoine Bosselut, Ronan Le Bras, Nouha Dziri, Xiang Li, Xuezhe Ma, Navonil Majumder, Debjit Paul, Soujanya Poriya, Simon Razniewski, Dan Roth, Keisuke Sakaguchi, Niket Tandon

Computational Social Science and Cultural Analytics
Somak Aditya, Abhijnan Chakraborty, Anjalia Elena Field, Andrew Halterman, Abram Handler, Kristen Johnson, Katherine Keith, Ashique KhudaBukhsh, Srijan Kumar, Sumeet Kumar, Roy Ka-Wei Lee, Animesh Mukherjee, Usman Naseem, Brendan O’Connor, Alice Oh, Eugenia Rho, Koustuv Saha, H. Andrew Schwartz, Qinlan Shen, Rob Voigt, Justine Zhang

Dialogue and Interactive Systems
Malihe Alikhani, Paul Crook, Devamanyu Hazarika, Simon Keizer, Bill Yuchen Lin, Odhisattwa Majumder, Alexandros Papangelis, Xiaojun Quan, Samira Shaikh, Lei Shu, Anthony Sicilia, Fanghua Ye, Zhou Yu, Yi Zhang

Discourse and Pragmatics
Christian Hardmeier, Yufang Hou, Feng Jiang, Shafiq Joty, Murathan Kurfaß, Ekaterina Lapshinova-Koltunski, Junyi Jessy Li, Yang Janet Liu, Sharid Loäigica, Vincent Ng, Tatjana Scheffler, Sebastian Schuster, Manfred Stede, Juntao Yu, Amir Zeldes

Efficient Methods for NLP
Cody Blakeney, Patrick H. Chen, Tim Dettmers, Beyza Ermis, Utku Evei, Sara Hooker, Yuhang Li, Bradley McDanel, Vishvak Murahari, Jonas Pfeiffer, Edoardo Ponti, Mohammad Rostami, Andreas Rücklé, Tal Schuster, Roy Schwartz, Dustin Wright, Mengzhou Xia, Ahmet Üstün

Ethics in NLP
Marta R. Costa-jussà, Kathleen Fraser, Lea Frermann, Dirk Hovy, Anne Lauscher, Margot Mieskes, Preslav Nakov, Debora Nozza, Chan Young Park, Sherif Saad, Zeerak Talat, Jieyu Zhao

Human-Centered NLP
Cecilia O. Alm, Jeffrey P. Bigham, Su Lin Blodgett, Kianté Brantley, Marine Carpuat, Hal Daumé III, Kenny Davila, Shi Feng, Kristina Gligoric, Saad Hassan, Michael A. Madaio, Khanh Xuan Nguyen, Weiyan Shi, Alison Smith, Sherry Wu, Tongshuang Wu, Diyi Yang
Information Extraction

Jun Araki, Yixin Cao, Muhao Chen, Hao Cheng, Christos Christodoulopoulos, Asif Ekbal, Hao Fei, Debasis Ganguly, Xu Han, Ruihong Huang, Fei Li, Hongyu Lin, Kang Liu, Tingwen Liu, Zhiyuan Liu, Wei Lu, Bhaskar Mitra, Sudip Naksar, Thien Huu Nguyen, Qiang Ning, Ankur Padia, Jay Pujara, Alan Ritter, Mrinmaya Sachan, Yu Su, Scott Yih, Wenpeng Yin, Yue Zhang

Information Retrieval and Text Mining

Maria Antoniak, Dallas Card, Fabio Crestani, Wenqi Fan, Yi Fang, Fuli Feng, Lifu Huang, Yiqun Liu, David Mimno, Jian-Yun Nie, Gabriella Pasi, Denis Peskoff, Lynda Lechani Tamine, Jinggang Wang, Qifan Wang, Zenglin Xu

Interpretability, Interactivity and Analysis of Models for NLP


Language Grounding to Vision, Robotics and Beyond

Mohit Bansal, Yonatan Bisk, Valts Blukis, Daniel Fried, Zhe Gan, Aniruddha Kembhavi, Jie Lei, Paul Pu Liang, Fangyu Liu, Roma Patel, Jivko Sinapov, Alane Suhr, Hao Tan, Xin Eric Wang, Zirui Wang

Language Modeling and Analysis of Language Models

Jan Buys, Kevin Clark, Sandipan Dandipat, Kumar A Dubey, Orhan Firat, Manish Gupta, Hany Hassan, Harsh Jhamtani, Melvin Johnson, Mandar Joshi, Urvashi Khandelwal, Lingpeng Kong, Ni Lao, Moontae Lee, Bing Liu, Peter J Liu, Qiang Liu, Colin Raffel, Naomi Saphra, Emma Strubell, Huan Sun, Partha Pratim Talukdar, Lijun Wu, Chunting Zhou

Linguistic Theories, Cognitive Modeling, and Psycholinguistics

Carolyn Jane Anderson, Damien Blasi, Ritesh Kumar, Yunfei Long, Tiago Pimentel, Géraldine Walther, Philipp Wicke

Machine Learning for NLP

Heike Adel, Mikhail Burtsev, Giuseppe Castellucci, Kai-Wei Chang, Sourish Chaudhuri, Wenhu
Chen, Arman Cohan, Danilo Croce, Julian Martin Eisenschlos, Francis Ferraro, Matthias Gallé, Philip John Gorinski, Tatsunori Hashimoto, Junxian He, Ricardo Henao, Jack Hessel, Estevam Hruschka, Dongyeop Kang, Pei Ke, Yoon Kim, Jay Yoon Lee, Irene Li, Lei Li, Zemin Liu, Ashutosh Modi, Thanh Tam Nguyen, Giannis Nikolentzos, Raphael Olivier, Ankur P Parikh, Danish Pruthi, Bhiksha Raj, Pasquale Restaino, Freda H. Shi, Jun Suzuki, Swabha Swayamdipta, Hao Tang, Pat Verga, William Wang, Taro Watanabe, Junyu Xuan, Kaisheng Yao, Se-Young Yun, Ningyu Zhang

Machine Translation
Loic Barrault, Rachel Bawden, Laurent Besacier, Alexandra Birch, Boxing Chen, Colin Cherry, Angela Fan, Marcello Federico, Mark Fishel, George Foster, Alexander Fraser, Markus Freitag, Thanh-Le Ha, Barry Haddow, Gholamreza Haffari, Lifeng Han, Shujian Huang, Philippe Langlais, Jindrich Libovicky, Qun Liu, Xuebo Liu, Lili Mou, Masaaki Nagata, Stephan Peitz, Maja Popovic, Anoop Sarkar, Holger Schwenk, Rico Sennrich, Felix Stahlberg, Zhaopeng Tu, David Vilar, Joern Wuebker, Tong Xiao, François Yvon

Multilinguality and Linguistic Diversity
David Ifeoluwa Adelani, Priyanka Agrawal, Antonios Anastopoulos, Mikel Artetxe, A. Seza Doğruöz, Yoshinari Fujinuma, Dan Garrette, Constantine Lignos, Manuel Mager, Benjamin Müller, Shruti Rijhwani, Sunayana Sitaram, Xinyi Wang, Guillaume Wisniewki, Zheng-Xin Yon, Marcos Zamipier

Natural Language Generation

NLP Applications
Kolawole Adebayo, Fernando Alva-Manchego, Meriem Beloucif, Daniel M. Bikel, Edward Choi, Manuel Rafael Ciosici, Colin Clement, Tirthankar Ghosal, Zhijiang Guo, Andrea Horbach, Daiki Kimura, Ekaterina Kochmar, Yanchi Liu, Nitin Madnani, Nedjma Ousidhoum, Natalie Parde, Lis Pereira, Daniel Preotiuc-Pietro, Marek Rei, Alla Rozovskaya, Michael Sejr Schlichtkrull, Kevin Small, Giancarlo Sperli, Anaïs Tack, James Thorne, Subhashini Venugopalan, V.G. Vinod Vydishwaran, Xuan Wang, Qianqian Xie, Ruiyong Xu, Victoria Yaneva, Min Yang, Torsten Zesch, Taolin Zhang, Xin Zhao, Arkaitz Zubiaga

Phonology, Morphology, and Word Segmentation
Claudia Borg, Ryan Cotterell, Yugo Murawaki, Emily Tucker Prud’hommeaux, Brian Roark, Kai-rit Sirts, Ekaterina Vylomova
Question Answering
Roee Aharoni, Akari Asai, Iz Beltagy, Pradeep Dasigi, Simone Filice, Siddhant Garg, Hannaneh Hajishirzi, Jonathan Herzig, Robin Jia, Daniel Khashabi, Tom Kwiatkowski, Ivan-no Lauriola, Mengwen Liu, Sewon Min, Alessandro Moschitti, Masud Moshtaghi, Salvatore Romeo, Min Joon Seo, Olga Uryupina, Thuy Vu

Resources and Evaluation
Firoj Alam, Federico Boschetti, Emily Chen, Vera Demberg, Rotem Dror, Iria de-Dios-Flores, Angelo Mario del Grosso, Yannick Estève, Nazli Goharian, Valeria Quochi, Sophie Rosset, Benoit Sagot, Yves Scherrer, Claudia Soria, Sajad Sotudeh, Daan Van Esch, Wajdi Zaghouani

Semantics: Lexical, Sentence level, Document Level, Textual Inference, etc.
Alan Akbik, Marianna Apidianaki, Valerio Basile, Eduardo Blanco, Elena Cabrio, Emmanuele Chersoni, Cristina España i Bonet, Allyson Ettinger, Goran Glavas, Andrey Kutuzov, Sobha Lathia Devi, Alessandro Lenci, Nafise Sadat Moosavi, Roser Morante, Constantin Orasan, Mohammad Taher Pilehvar, Yuval Pinter, Lidia Pivovarova, Alessandro Raganato, Paul Rayson, Steven Schockaert, Ivan Vulić, Loneke van der Plas, Gijs Wijnholds, Wei Xu, Sho Yokoi

Sentiment Analysis, Stylistic Analysis, and Argument Mining
Nikolaos Aletras, Ehsaneddin Asgari, Isabelle Augenstein, Nathaniel Blanchard, David Jurgens, Elena Kochkina, Lun-Wei Ku, Ivan Vladimir Meza Ruíz, Saif Mohammad, Joonsuk Park, Mohammad Taher Pilehvar, Ehsan Shareghi, Shabnam Tafreshi, Eva Maria Vecchi, Serena Villata, Hao Wang, Shuai Wang, Yadollah Yaghoobzadeh

Speech and Multimodality
Ahmed Ali, Shammur Absar Chowdhury, David Harwath, Takaaki Horiat, Wei-Ning Hsu, Ryo Ishii, Herman Kamper, Suyoun Kim, Ondrej Klejch, Gakuto Kurata, Sheng Li, Erfan Loweimi, Soumi Maiti, Florian Metze, Shrutí Palaskar, Leibny Paola Garcia Perera, Murat Saracılı, Suwon Shon, Rita Singh, Katsuhito Sudoh, Shinji Watanabe, Matthew Wiesner, Koichiro Yoshino

Summarization
Reinald Kim Amplayo, Florian Boudin, Yue Dong, Yang Gao, Tanya Goyal, Hiroaki Hayashi, Sathish Reddy Indurthi, Jey Han Lau, Manling Li, Marina Litvak, Fei Liu, Annie Louis, Joshua Maynez, Aiala Rosá, Horacio Saggion, Xiaodan Zhu

Syntax, Parsing and their Applications
Agnieszka Falenska, Liang Huang, Zhenghua Li, Haitao Mi, Joakim Nivre, Kenji Sagae, Kewei Tu, David Vilares, Rob van der Goot, Meishan Zhang
Theme Track: Large Language Models and the Future of NLP

Primary Reviewers

Ann Clifton, Anne Cocos, Davide Colla, Pedro Colon-Hernandez, Andrei Catalin Coman, Anna Corazza, Francesco Corcoglioniti, Camille Couturier, Benoit Crabbé, Mathias Creutz, Liam Crippwell, James Cross, Maxwell Crouse, Gangu Cui, Leyang Cui, Peng Cui, Shaobo Cui, Shiyou Cui, Wenjia Cui, Yiming Cui, Rossana Cunha


Alexander Fabbri, Marzieh Fadaee, Hossein Rajab Faghihi, Fahim Faisal, Ge Fan, Jungwei Fan, Run-Ze Fan, Yao-Chung Fan, Yue Fan, Biaoyan Fang, Jinyuan Fang, Liri Fang, Qingkai Fang, Tianqing Fang, Wanlong Fang, Wei Fang, Xiang Fang, Yimai Fang, Yuwei Fang, Hossein Fani, Ibrahim Abu Farha, Effat Farhana, Nawshad Farruque, Amirhossein Farzam, Amany Fashwan, Jean-Philippe Fauconnier, Adam Faulkner, Amir Feder, Zichu Fei, Aarash Feizi, Nils Feldhus, Anna Feldman, Sergey Feldman, Virginia K. Feltner, Jia Hui Feng, Jianzhou Feng, Jiazhan Feng, Jingrong Feng, Shubing Feng, Shutong Feng, Steven Y. Feng, Wei Xi Feng, Xiaochong Feng, Yalin Feng, Yi Feng, Yu Feng, Yujie Feng, Yunhe Feng, Zhangyu Feng, Paulo Fernandes, Nigel Fernandez, Raquel Fernández, Elisa Ferracane, Javier Ferrando, Rafael Ferreira, Manuel Vilares Ferro, Elisabetta Fersini, Besnik Fetahu, Alejandro Figueroa, Marcelo Finger, Matthew Finlayson, Maujama Firdaus, Tim Fischer, Margaret M. Fleck, Michael Flor, Marco Fonseca, Jennifer Foster, Nathan Fradet, Marc Franco-Salvador, Thomas Français, Flavius Frasincar, Dayne Freitag, Francesca Frontini, Deqing Fu, Haomin Fu, Jie Fu, Lisheng Fu, Peng Fu, Quchen Fu, Tingchen Fu, Tsu-Jui Fu, Xingyu Fu, Fumiyo Fukumoto, Richard Futrell, Michael Färber

Matteo Gabbri, Kata Gabor, Marco Gaido, Amit Gajbhiye, Lukas Galke, Leilei Gan, Yanglei Gan, Yujian Gan, Sudeep Gandhe, Ashwinkumar Ganesh, Ananya Ganesh, Govind Krishnan Gangadhar, Achyutarama R Ganti, William Gantt, Chang Gao, Chongyang Gao, Ge Gao, Hon-
Zhongqiang Huang, Zijian Huang, Ziyang Huang, John S Hudzina, Binyuan Hui, Wenyang Hui, Zheng Hui, Chia-Chien Hung, Kyunghoon Hur, Tin Van Huynh, Dae Yon Huang, Truong Son Hy, Dongmin Hyun, Katharina Hämmerl, Ali Hürriyetoğlu

Ignacio Iacobacci, Adrian Iftene, Ryu lida, Nikolai Ilinykh, Kenji Imamura, Joseph Marvin Imperial, Hirofumi Inaguma, Koji Inoue, Takashi Inui, Tatsuya Ishigaki, Shotaro Ishihara, Etsuko Ishii, Aminul Islam, Tunazzina Islam, Hayate Iso, Masaru Isonuma, Takumi Ito, Hamish Ivison, Tomoya Iwakura, Ran Iwamoto, Kenichi Iwatsuki, Arun Iyer, Roshni Iyer, Vivek Iyer, Peter Izsak


Song, Ran Song, Rui Song, Yifan Song, Zhenqiao Song, Nikita Soni, Sandeep Soni, Alexey Sorokin, Dmitry Sotnikov, Anna Sotnikova, Najib Sotudeh, Marlo Souza, Gerasimos Spanakis, Timo Spinde, Cesare Spinoso-Di Piano, Andreas Spitz, Makers Narsimhan Sreedhar, Arvind Krishna Sridhar, Sharath Nittur Sridhar, Mukund Srinath, Gokul Srinivasan, Balaji Vasan Srinivasan, Pranesh Srinivasan, Tejas Srinivasan, Aarohi Srivastava, Ankit Kumar Srivastava, Aseem Srivastava, Ieva Stalunaitė, Dominik Stammbach, Karolina Stanczak, Marija Stanojevic, Gabriel Stanovsky, David Stap, Katherine Stasaski, Julius Steen, Shane Steinert-Threlkeld, Georg Stemmer, Elias Stengel-Eškin, Andreas Stephan, Zachary Stine, Alessandro Stolfo, Shane Storkes, Tomek Strzalkowski, Phillip Benjamin Ströbel, Sara Stymne, Sebastian Stüker, Hsuan Su, Jinyan Su, Miqo Su, Qiag Su, Ruolin Su, Xiangdong Su, Xin Su, Yuefeng Su, Ying Su, Yixuan Su, Pedro Ortiz Suarez, Nishant Subramani, Prasanna Lakshmi Subramanyam, Smitha Muthya Sudheendra, Hiroaki Sugiyama, Kazunari Sugiyama, Yoshi Suhara, Xuhui Sui, Elior Sulem, Md Arafat Sultan, Albert Sun, Changzhi Sun, Chengjie Sun, Chenkai Sun, Fei Sun, Guanzhi Sun, Haipeng Sun, Hao Sun, Hao Sun, Hong Sun, Jian Sun, Kaiser Sun, Kun Sun, Le Sun, Ming Sun, Peijie Sun, Qiushi Sun, Renliang Sun, Rui Sun, Shichao Sun, Simeng Sun, Tianxiang Sun, Weiwei Sun, Weixuan Sun, Xiaoyang Sun, Yi Sun, Yuta Sun, Zengkui Sun, Zequn Sun, Zewei Sun, Zhiqing Sun, Zhoujian Sun, Mujeen Sung, Yoo Yeon Sung, Hanna Suominen, Marek Suppa, Abhijit Suresh, Allmic Pradhap Singh Susaiyah, Andrias Susanto, Lintang Sutawika, Benjamin Suter, Mircuz Sun, Sarathkrisna Swaminathan, Sandesh Swamy, Munira Syed, Stan Szpakowicz, Maria Sänger, Joan Andreu Sánchez, Ricardo Muñoz Sánchez, Victor M. Sánchez-Cartagena, Michal Štefánik

Santosh T.Y.S.S, Jeniya Tabassum, Oyvind Tafjord, Chang-Yu Tai, Dima Taji, Yu Takagi, Sho Takase, Zeerak Talat, George Tambouratzis, Alec Tamchyna, Chao-Hong Tan, Haochen Tan, Liling Tan, Minghuan Tan, Qingyu Tan, Xingwei Tan, Ryota Tanaka, Gongbo Tang, Haoyu Tang, Hui Tang, Liyan Tang, Ruixiang Tang, Shuai Tang, Tianyi Tang, Wei Tang, Xiangru Tang, Xue-mei Tang, Xunzhu Tang, Yihong Tang, Zheng Tang, Zhiwen Tang, Zineng Tang, Guanhong Tao, Mingyu Tao, Wei Tao, Sandeep Tata, Marta Tatu, Selma Tekir, Serra Sinem Tekiroğlu, Eric S. Tellez, Irina Temnikova, Zhiyang Teng, Davide Testa, Alberto Testoni, Martin Teuffenbach, Katherine Thai, Khushboo Thaker, Urmish Thakker, Himanshu Thakur, Surendrabikram Thapa, Avijit Thawani, Anton Frederik Thielmann, Camillo Thorne, Ron Tan, Xuetao Tian, Yijun Tian, Yuanhe Tian, Yufei Tian, Yuhang Tian, Zhiliang Tian, Jörg Tiedemann, Zoran Tiganj, Abhishek Tiwari, Nidhi Tiwari, Prayag Tiwari, Soham Dinesh Tiwari, Hung Quoc To, Amalia Todirascu, Sebastian-Antonio Toma, Nadi Tomeh, Nicholas Tomlin, Cagri Toraman, Shubham Toshniwal, Samia Touileb, Yannick Toussaint, Benjamin Towl, Khanh Quoc Tran, Khiem Vinh Tran, Lucas Vinh Tran, Son Quoc Tran, Thi Hong Hanh Tran, Long Hai Trieu, Bayu Distiawan Trisedya, Harsh Trivedi, Sergey Troshin, Adam Tsakalidis, Dimitrios Tsarapatsanis, Bo-Hsiang Tseng, Ioannis Tsiamas, Eleftheria Tsipidi, Masaaki Tsuchida, Geng Tu, Jingxuan Tu, Yunbin Tu, Yi-Lin Tuan, Marco Turchi

Irina Ualiyeva, Rutuja Ubale, Ana Sabina Uban, Solomon Ubani, Oseremen Oscar Uduehi, Adrian Ulges, Eddie L. Ungless, Apoorva Upadhyaya, Gorka Urbizu, Ashok Urlana, Asahi Ushio, Saiteja Utpala


Emily Öhman, Lilja Ovrelid, Tolúlopé Ògúnrèmí
Keynote Talk: Human-Centric Natural Language Processing

Jong C. Park
Korea Advanced Institute of Science and Technology

Friday, December 8 – Time: 09:30 - 10:30 – Room: East & Center

Abstract: In keeping up with the three pillars of natural language processing, ever-changing principles and techniques, emerging domains of application and related resources, and people with needs for language support, our research team has paid attention to the third pillar with a special focus on human-centric diversity and minority issues, including children, experts in biology and medicine, deaf people, and people with different challenges to language use. In this talk, I wish to share our achievements along the way, recently leading to a flurry of pleasant results that make use of deep learning techniques and large language models. I conclude the talk with a cautious prognosis about what might lie beyond large language models that loom over much of what we do at the moment and may hamper diversity to a worrisome degree and even humanity in hindsight.

Bio: Jong Park received his BE and MSE degrees from Seoul National University and PhD degree from the University of Pennsylvania, Philadelphia. He has been working as Assistant, Associate and Full Professor at Korea Advanced Institute of Science and Technology (KAIST) since 1998. He is one of the early researchers on BioNLP, applying NLP techniques to biology and medicine. His research team at KAIST has also been working broadly on identifying emotion from text, turning spoken language into visual animation and sign language, identifying mental health issues such as mild-cognitive impairment (MCI) and clinical depression from natural language utterances, detecting abusive language, and, more recently, credibility assessment and bidirectional sign language processing. His team has received Outstanding Paper Award at ACL 2023 for the work on the Tigrinya language. He is serving as founding Editor-in-Chief of Journal of Computing Science and Engineering (JCSE) since 2007, President of Asian Federation of Natural Language Processing (AFNLP) during 2022–2024, and General Chair of IJCNLP-AAACL 2023.
Keynote Talk: From Speech to Emotion to Mood: Mental Health Modeling in Real-World Environments

Emily Mower Provost
University of Michigan

Saturday, December 9 – Time: 14:30 - 15:30 – Room: East & Center

Abstract: Emotions provide critical cues into our health and wellbeing. This is of particular importance in the context of mental health, where changes in emotion may signify changes in symptom severity. However, information about emotion and how it varies over time is often accessible only using survey methodology (e.g., ecological momentary assessment, EMA), which can become burdensome to participants over time. Automated speech emotion recognition systems could provide an alternative, providing quantitative measures of emotion using acoustic data captured passively from a consented individual’s environment. However, speech emotion recognition systems often falter when presented with data collected from unconstrained natural environments due to issues with robustness, generalizability, and invalid assumptions. In this talk, I will discuss our journey in speech-centric mental health modeling, explaining whether, how, and when emotion recognition can be applied to natural unconstrained speech data to measure changes in mental health symptom severity.

Bio: Emily Mower Provost is a Professor in Computer Science and Engineering at the University of Michigan. She received her Ph.D. in Electrical Engineering from the University of Southern California (USC), Los Angeles, CA in 2010. She is a Toyota Faculty Scholar (2020) and has been awarded a National Science Foundation CAREER Award (2017), the Oscar Stern Award for Depression Research (2015), a National Science Foundation Graduate Research Fellowship (2004-2007). She is an Associate Editor for IEEE Transactions on Affective Computing and the IEEE Open Journal of Signal Processing. She has also served as Associate Editor for Computer Speech and Language and ACM Transactions on Multimedia. She has received best paper awards or finalist nominations for Interspeech 2008, ACM Multimedia 2014, ICMI 2016, and IEEE Transactions on Affective Computing. Among other organizational duties, she has been Program Chair for ACII (2017, 2021), ICMI (2016, 2018). Her research interests are in human-centered speech and video processing, multimodal interfaces design, and speech-based assistive technology. The goals of her research are motivated by the complexities of the perception and expression of human behavior.
Keynote Talk: Academic NLP research in the Age of LLMs: Nothing but blue skies!

Christopher D. Manning
Stanford University

Sunday, December 10 – Time: 14:00 - 15:00 – Room: East & Center

Abstract: There has been a certain amount of handwringing by students – and faculty – about the prospects for academic research in the Age of Large Language Models (LLMs). Now, there is a need for universities and governments to do more to support computational research. And it is indeed the case that any major inflection point in a research field does change the best questions to concentrate your time on. But a reorientation of research is healthy, coming off a decade when students in academia had often drifted into playing the Kaggle game rather than thinking hard about fundamental ideas. A time of transformation is principally a time of opportunity: All sorts of new and exciting research questions, either not explored or only explored in very different contexts, come to the fore. These questions provide compelling opportunities for fresh and exciting work. I see many examples of this sort of work already happening and encourage many more. I will illustrate with two pieces of work from my own students: the Backpack, which explores an alternative, more interpretable architecture than the Transformer (by John Hewitt and colleagues) and Direct Preference Optimization, which explores an alternative to the Proximal Policy Optimization normally used for steering LLMs in the Reinforcement Learning from Human Feedback (RLHF) phase (by Eric Mitchell and colleagues).

Bio: Christopher Manning is the inaugural Thomas M. Siebel Professor in Machine Learning in the Departments of Linguistics and Computer Science at Stanford University, Director of the Stanford Artificial Intelligence Laboratory (SAIL), and an Associate Director of the Stanford Institute for Human-Centered Artificial Intelligence (HAI). His research goal is computers that can intelligently process, understand, and generate human languages. Manning was an early leader in applying Deep Learning to Natural Language Processing (NLP), with well-known research on the GloVe model of word vectors, attention, machine translation, question answering, self-supervised model pre-training, tree-recursive neural networks, machine reasoning, dependency parsing, sentiment analysis, and summarization. He also focuses on computational linguistic approaches to parsing, natural language inference and multilingual language processing, including being a principal developer of Stanford Dependencies and Universal Dependencies. Manning has coauthored leading textbooks on statistical approaches to NLP (Manning and Schütze 1999) and information retrieval (Manning, Raghavan, and Schütze, 2008), as well as linguistic monographs on ergativity and complex predicates. His online CS224N Natural Language Processing with Deep Learning videos have been watched by hundreds of thousands of people. He is an ACM Fellow, a AAAI Fellow, and an ACL Fellow, and a Past President of the ACL (2015). His research has won ACL, Coling, EMNLP, and CHI Best Paper Awards, and an ACL Test of Time Award. He has a B.A. (Hons) from The Australian National University and a Ph.D. from Stanford in 1994, and an Honorary Doctorate from U. Amsterdam in 2023, and he held faculty positions at Carnegie Mellon University and the University of
Sydney before returning to Stanford. He is the founder of the Stanford NLP group (@stanfordnlp) and manages development of the Stanford CoreNLP and Stanza software.
Panel: Beyond Text: Inclusive Human Communication with Language Technology

Saturday, December 9 – Time: 16:00 - 17:00 – Room: East & Center

The scope of communication extends far beyond textual interactions. This panel converges experts from diverse fields, including sign language, multilingual conference interpretation, empathetic speech-based dialog systems, virtual reality, and embodied agents, who will discuss how advancements in language technology can help shape the future of inclusive communication, and at the same time what NLP can learn from various fields of communication.

Panelists:

Ms. Lourdes de Rioja, Freelance Conference Interpreter, Belgium
Lourdes de Rioja is a Spanish conference interpreter who has worked for the European Commission, the European Parliament, and the European Court of Justice. She is the author and producer of SCICtrain for the European Commission, editor and designer of False Friends online Dictionary, and co-creator of AIIC Conversations. Ms de Rioja has a Master’s degree in Political & Corporate Communication from The George Washington University & Universidad de Navarra, Spain, and a European Master’s degree in Conference Interpreting from Universidad de La Laguna, Spain. She is a member of the International Association of Conference Interpreters (AIIC) and has interpreted in various languages, including English, French, Danish, Swedish, and Catalan.

Dr. Abraham Glasser, Gallaudet University, USA
Dr. Abraham Glasser is an Assistant Professor in Science, Technology, Accessibility, Mathematics, and Public Health. He earned his undergraduate degree and his Ph.D. in Computing and Information Sciences from the Rochester Institute of Technology. Glasser has conducted research work for various organizations. His research interests include virtual and augmented reality, accessible technology for deaf and hard of hearing individuals – with a focus on automatic speech recognition, and he is passionate about making technology inclusive for people with disabilities. Dr. Glasser is also the reigning United States Deaf Chess Champion.

Prof Chengkuo Lee, National University, Singapore
Prof Chengkuo Lee is an Associate Professor at the Dept. of Electrical and Computer Eng. of National University of Singapore, Singapore with a research background in Systems and Precision Engineering. He received his M.S. degree in Materials Science and Engineering from National Tsing Hua University, Hsinchu, Taiwan, in 1991, and his Ph.D. degree in Precision Engineering from the University of Tokyo, Tokyo, Japan, in Jan. 1996. He has proposed augmented tactile-perception and haptic-feedback rings with multimodal sensing and feedback capabilities for Human-Machine interfaces for immersive interactions. He has also proposed a sign language recognition and communication system using a smart triboelectric glove, AI block, and the back-end VR interface. Prof. Lee has contributed more than 250 papers in peer-reviewed international journals and conferences, and 9 US patents in MEMS, NEMS, Nanophotonics and Nanotechnology fields.

Prof María Inés Torres, UPV/EHU, Spain
María Inés Torres is a Professor of Computer Science at the University of the Basque Country. She founded the Pattern Recognition and Speech Technology research group in 1990, which she has been leading
since then. She has conducted research related to speech technologies, including automatic speech recognition and understanding, language identification, machine translation, specific processing of Basque language, and acquisition and generation of language resources. She has published numerous papers in journals and international conferences and edited three books. Her current research interests focus on statistical approaches to deal with spoken dialog systems, being also interested in learning from human interaction to generate artificial interaction.

Prof Monojit Choudhury, MBZUAI, Abu Dhabi
Monojit Choudhury is a professor of NLP at MBZUAI, Abu Dhabi. He has a PhD and BTech in computer science and engineering from IIT Kharagpur. Prof Choudhury’s research interests span various aspects of natural language processing and cognitive sciences, including multilingual models, computational ethics, and AI and NLP for social good. Prior to joining MBZUAI, he was a principal researcher at Microsoft. He also serves as a professor of practice at Plaksha University and an adjunct faculty at IIIT Hyderabad. Prof Choudhury is the general chair of the Panini Linguistics Olympiad, which is the Indian national linguistics Olympiad for high school students and the founding co-chair of the Asia Pacific Linguistics Olympiad.
# Table of Contents

**IAG: Induction-Augmented Generation Framework for Answering Reasoning Questions**  
Zhebin Zhang, Xinyu Zhang, Yuanhang Ren, Saijiang Shi, Meng Han, Yongkang Wu, Ruofei Lai and Zhao Cao ................................................................. 1

**Absolute Position Embedding Learns Sinusoid-like Waves for Attention Based on Relative Position**  
Yuji Yamamoto and Takuya Matsuzaki .......................................................... 15

**Chinese Lexical Substitution: Dataset and Method**  
Jipeng Qiang, Kang Liu, Ying Li, Yun Li, Yi Zhu, Yun-Hao Yuan, Xiaocheng Hu and Xiaoye Ouyang ................................................................. 29

**Decoding the Silent Majority: Inducing Belief Augmented Social Graph with Large Language Model for Response Forecasting**  
Chenkai Sun, Jinning Li, Yi Fung, Hou Pong Chan, Tarek Abdelzaher, ChengXiang Zhai and Heng Ji ................................................................. 43

**Fine-grained Conversational Decoding via Isotropic and Proximal Search**  
Yuxuan Yao, Han Wu, Qiling Xu and Linqi Song ............................................ 58

**Holistic Inter-Annotator Agreement and Corpus Coherence Estimation in a Large-scale Multilingual Annotation Campaign**  
Nicolas Stefanovitch and Jakub Piskorski ................................................. 71

**PHD: Pixel-Based Language Modeling of Historical Documents**  
Nadav Borenstein, Phillip Rust, Desmond Elliott and Isabelle Augenstein ............... 87

**Primacy Effect of ChatGPT**  
Yiwei Wang, Yujun Cai, Muha Chen, Yuxuan Liang and Bryan Hooi ..................... 108

**Evaluating the Rationale Understanding of Critical Reasoning in Logical Reading Comprehension**  
Akira Kawabata and Saku Sugawara .......................................................... 116

**Evaluating and Modeling Attribution for Cross-Lingual Question Answering**  
Benjamin Muller, John Frederick Wieting, Jonathan H. Clark, Tom Kwiatkowski, Sebastian Ruder, Livio Baldini Soares, Roe Josang, Jonathan Herzig and Xinyi Wang ......................... 144

**Better Quality Pre-training Data and T5 Models for African Languages**  
Akintunde Oladipo, Mofetoluwa Adeyemi, Orewaogheni Achia, Abraham Toluwa Owodunni, Odunayo Ogundepo, David Ifeoluwa Adelani and Jimmy Lin .................. 158

**Sparse Universal Transformer**  
Shawn Tan, Yikang Shen, Zhenfang Chen, Aaron Courville and Chuang Gan ............. 169

**Theory of Mind for Multi-Agent Collaboration via Large Language Models**  
Huao Li, Yu Quan Chong, Simon Stepputts, Joseph Campbell, Dana Hughes, Charles Michael Lewis and Katia P. Sycara .................................................. 180

**Establishing Trustworthiness: Rethinking Tasks and Model Evaluation**  
Robert Litschko, Max Müller-Eberstein, Rob van der Goot, Leon Weber-Genzel and Barbara Plank 193
Let’s Think Frame by Frame with VIP: A Video Infilling and Prediction Dataset for Evaluating Video Chain-of-Thought
Vaishnavi Himakunthala, Andy Ouyang, Daniel Philip Rose, Ryan He, Alex Mei, Yujie Lu, Chinmay Sonar, Michael Saxon and William Yang Wang ................................................................. 204

GPTAraEval: A Comprehensive Evaluation of ChatGPT on Arabic NLP
Md Tawkat Islam Khondaker, Abdul Waheed, El Moatez Billah Nagoudi and Muhammad Abdul-Mageed ............................................................... 220

Dual-Channel Span for Aspect Sentiment Triplet Extraction
Pan Li, Ping Li and Kai Zhang ................................................................. 248

Cultural Concept Adaptation on Multimodal Reasoning
Zhi Li and Yin Zhang ........................................................................ 262

Understanding Compositional Data Augmentation in Typologically Diverse Morphological Inflection
Farhan Samir and Miikka Silfverberg ..................................................... 277

Evaluating Object Hallucination in Large Vision-Language Models
Yifan Li, Yifan Du, Kun Zhou, Jinpeng Wang, Xin Zhao and Ji-Rong Wen ................................................................. 292

Event Ontology Completion with Hierarchical Structure Evolution Networks
Pengfei Cao, Yupu Hao, Yubo Chen, Kang Liu, Jiexin Xu, Huaijun Li, Xiaoqian Jiang and Jun Zhao ................................................................. 306

Parameter-efficient Tuning for Large Language Model without Calculating Its Gradients
Feihu Jin, Jiajun Zhang and Chengqing Zong .......................................... 321

Discourse Structures Guided Fine-grained Propaganda Identification
Yuanyuan Lei and Ruihong Huang ......................................................... 331

CompoundPiece: Evaluating and Improving Decompounding Performance of Language Models
Benjamin Minixhofer, Jonas Pfeiffer and Ivan Vulić ................................ 343

Improving Image Captioning via Predicting Structured Concepts
Ting Wang, Weidong Chen, Yuanhe Tian, Yan Song and Zhendong Mao ......................................................................................... 360

GATTITOS: Using a New Multilingual Lexicon for Low-resource Machine Translation
Alexander Jones, Isaac Rayburn Caswell, Orhan Firat and Ishank Saxena ......................................................................................... 371

Continually Improving Extractive QA via Human Feedback
Ge Gao, Hung-Ting Chen, Yoav Artzi and Eunsol Choi ................................................................. 406

Using Interpretation Methods for Model Enhancement
Zhuo Chen, Chengyue Jiang and Kewei Tu ........................................... 424

An Expression Tree Decoding Strategy for Mathematical Equation Generation
Wenqi Zhang, Yongliang Shen, Qingpeng Nong, Zeqi Tan, Yanna Ma and Weiming Lu ................................................................. 439

Bootstrapping Small & High Performance Language Models with Unmasking-Removal Training Policy
Yahan Yang, Elior Sulem, Insup Lee and Dan Roth ........................................ 457

Diversity Enhanced Narrative Question Generation for Storybooks
Hokeun Yoon and JinYeong Bak ............................................................. 465

Debiasing Made State-of-the-art: Revisiting the Simple Seed-based Weak Supervision for Text Classification
Chengyu Dong, Zihan Wang and Jingbo Shang ........................................ 483

xxix
How to Enhance Causal Discrimination of Utterances: A Case on Affective Reasoning
Hang Chen, Xinyu Yang, Jing Luo and Wenjing Zhu .................................................. 494

Compressing and Debiasing Vision-Language Pre-Trained Models for Visual Question Answering
Qingyi Si, Yuanxin Liu, Zheng Lin, Peng Fu, Yanan Cao and Weiping Wang ............... 513

Selectively Answering Ambiguous Questions
Jeremy R. Cole, Michael JQ Zhang, Daniel Gillick, Julian Martin Eisenschlos, Bhuwan Dhingra and Jacob Eisenstein ................................................................. 530

Temporal Knowledge Graph Forecasting Without Knowledge Using In-Context Learning
Dong-Ho Lee, Kian Ahrabian, Woojeong Jin, Fred Morstatter and Jay Pujara ............. 544

Knowledge Graph Compression Enhances Diverse Commonsense Generation
EunJeong Hwang, Veronika Thost, Vered Shwartz and Tengfei Ma ......................... 558

Pragmatic Reasoning Unlocks Quantifier Semantics for Foundation Models
Yiyuan Li, Rakesh R Menon, Sayan Ghosh and Shashank Srivastava .................... 573

LLM-FP4: 4-Bit Floating-Point Quantized Transformers
Shih-yang Liu, Zechun Liu, Xijie Huang, Pingcheng Dong and Kwang-Ting Cheng .......... 592

Improving Biomedical Abstractive Summarisation with Knowledge Aggregation from Citation Papers
Chen Tang, Shun Wang, Tomas Goldsack and Chenghua Lin .................................. 606

Explanation Selection Using Unlabeled Data for Chain-of-Thought Prompting
Xi Ye and Greg Durrett .......................................................... 619

HalOmi: A Manually Annotated Benchmark for Multilingual Hallucination and Omission Detection in Machine Translation
David Dale, Elena Voita, Janice Lam, Prangthip Hansanti, Christophe Ropers, Elahe Kalbassi, Cynthia Gao, Loic Barrault and Marta R. Costa-jussà ........................................ 638

Gradient-based Gradual Pruning for Language-Specific Multilingual Neural Machine Translation
Dan He, Minh-Quang Pham, Thanh-Le Ha and Marco Turchi ................................ 654

LLM-powered Data Augmentation for Enhanced Cross-lingual Performance
Chenxi Whitehouse, Monojit Choudhury and Alham Fikri Aji ................................. 671

Prompt-based Logical Semantics Enhancement for Implicit Discourse Relation Recognition
Chenxu Wang, Ping Jian and Mu Huang ......................................................... 687

VLIS: Unimodal Language Models Guide Multimodal Language Generation
Jiwan Chung and Youngjae Yu ......................................................... 700

Conceptual structure coheres in human cognition but not in large language models
Siddharth Suresh, Kushin Mukherjee, Xizheng Yu, Wei-Chun Huang, Lisa Padua and Timothy T. Rogers ................................................................. 722

Towards LLM-driven Dialogue State Tracking
Yujie Feng, Zexin Lu, Bo Liu, Liming Zhan and Xiao-Ming Wu .......................... 739

Learning Language-guided Adaptive Hyper-modality Representation for Multimodal Sentiment Analysis
Haoyu Zhang, Yu Wang, Guanghao Yin, Kejun Liu, Yuanyuan Liu and Tianshu Yu ....... 756
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predict the Future from the Past? On the Temporal Data Distribution</td>
<td>Yue Guo, Chenxi Hu and Yi Yang</td>
<td>1029</td>
</tr>
<tr>
<td>Shift in Financial Sentiment Classifications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Look-back Decoding for Open-Ended Text Generation</td>
<td>Nan Xu, Chunting Zhou, Asli Celikyilmaz and Xuezhe Ma</td>
<td>1039</td>
</tr>
<tr>
<td>Large Language Models Can Self-Improve</td>
<td>Jiaxin Huang, Shixiang Shane Gu, Le Hou, Yuexin Wu, Xuezhi Wang, Hongkun Yu and Jiawei Han</td>
<td>1051</td>
</tr>
<tr>
<td>CodeT5+: Open Code Large Language Models for Code Understanding and Generation</td>
<td>Yue Wang, Hung Le, Akhilesh Deepak Gotmare, Nghi D. Q. Bui, Junnan Li and Steven Hoi</td>
<td>1069</td>
</tr>
<tr>
<td>Structural generalization in COGS: Supertagging is (almost) all you need</td>
<td>Alban Petit, Caio Filippo Corro and François Yvon</td>
<td>1089</td>
</tr>
<tr>
<td>BioT5: Enriching Cross-modal Integration in Biology with Chemical Knowledge and Natural Language Associations</td>
<td>Qizhi Pei, Wei Zhang, Jinhua Zhu, Kehan Wu, Kaiyuan Gao, Lijun Wu, Yingce Xia and Rui Yan</td>
<td>1102</td>
</tr>
<tr>
<td>Hyperpolyglot LLMs: Cross-Lingual Interpretability in Token Embeddings</td>
<td>Andrea W Wen-Yi and David Mimno</td>
<td>1124</td>
</tr>
<tr>
<td>Target-oriented Proactive Dialogue Systems with Personalization: Problem Formulation and Dataset Curation</td>
<td>Jian Wang, Yi Cheng, Dongding Lin, Chak Tou Leong and Wenjie Li</td>
<td>1132</td>
</tr>
<tr>
<td>SeqXGPT: Sentence-Level AI-Generated Text Detection</td>
<td>Pengyu Wang, Linyang Li, Ke Ren, Botian Jiang, Dong Zhang and Xipeng Qiu</td>
<td>1144</td>
</tr>
<tr>
<td>QTSumm: Query-Focused Summarization over Tabular Data</td>
<td>Yilun Zhao, Zhenting Qi, Linyong Nan, Boyu Mi, Yixin Liu, Weijin Zou, Simeng Han, Ruizhe Chen, Xiangru Tang, Yumo Xu, Dragomir Radev and Arman Cohan</td>
<td>1157</td>
</tr>
<tr>
<td>From Wrong To Right: A Recursive Approach Towards Vision-Language Explanation</td>
<td>Jiaxin Ge, Sanjay Subramanian, Trevor Darrell and Boyi Li</td>
<td>1173</td>
</tr>
<tr>
<td>‘Don’t Get Too Technical with Me’: A Discourse Structure-Based Framework for Automatic Science Journalism</td>
<td>Ronald Cardenas, Bingsheng Yao, Dakuo Wang and Yufang Hou</td>
<td>1186</td>
</tr>
<tr>
<td>Penalty Decoding: Well Suppress the Self-Reinforcement Effect in Open-Ended Text Generation</td>
<td>Wenhong Zhu, Hongkun Hao and Rui Wang</td>
<td>1218</td>
</tr>
<tr>
<td>Towards Robust Pruning: An Adaptive Knowledge-Retention Pruning Strategy for Language Models</td>
<td>Jianwei Li, Qi Lei, Wei Cheng and Dongkuan Xu</td>
<td>1229</td>
</tr>
<tr>
<td>Clinical Contradiction Detection</td>
<td>Dave Makhervaks, Plia Gillis and Kira Radinsky</td>
<td>1248</td>
</tr>
</tbody>
</table>
Vera: A General-Purpose Plausibility Estimation Model for Commonsense Statements
Jiacheng Liu, Wenya Wang, Dianzhuo Wang, Noah A. Smith, Yejin Choi and Hannaneh Hajishirzi

Text-Transport: Toward Learning Causal Effects of Natural Language
Victoria Lin, Louis-Philippe Morency and Eli Ben-Michael

How Does Generative Retrieval Scale to Millions of Passages?
Ronak Pradeep, Kai Hui, Jai Gupta, Adam D Lelkes, Honglei Zhuang, Jimmy Lin, Donald Metzler and Vinh Q. Tran

Unveiling the Implicit Toxicity in Large Language Models
Jiaxin Wen, Pei Ke, Hao Sun, Zhexin Zhang, Chengfei Li, Jinfeng Bai and Minlie Huang

Is ChatGPT a General-Purpose Natural Language Processing Task Solver?
Chengwei Qin, Aston Zhang, Zhuosheng Zhang, Jiaao Chen, Michihiro Yasunaga and Diyi Yang

Length is a Curse and a Blessing for Document-level Semantics
Chenghao Xiao, Yizhi Li, G Thomas Hudson, Chenghua Lin and Noura Al Moubayed

ALCUNA: Large Language Models Meet New Knowledge
Xunjian Yin, Baizhou Huang and Xiaojun Wan

Location-Aware Visual Question Generation with Lightweight Models
Nicholas Collin Suwono, Justin Chen, Tun Min Hung, Ting-Hao Kenneth Huang, I-Bin Liao, Yung-Hui Li, Lun-Wei Ku and Shao-Hua Sun

MemeCap: A Dataset for Captioning and Interpreting Memes
EunJeong Hwang and Vered Shwartz

Where to start? Analyzing the potential value of intermediate models
Leshem Choshen, Elad Venezian, Shachar Don-Yehiya, Noam Slonim and Yoav Katz

Transcending Scaling Laws with 0.1% Extra Compute
Yi Tay, Jason Wei, Hyung Won Chung, Vinh Q. Tran, David So, Siamak Shakeri, Xavier Garcia, Steven Zheng, Jinfeng Rao, Aakanksha Chowdhery, Denny Zhou, Donald Metzler, Slav Petrov, Neil Houlsby, Quoc V Le and Mostafa Dehghani

CoAnnotating: Uncertainty-Guided Work Allocation between Human and Large Language Models for Data Annotation
Minzhi Li, Taiwei Shi, Caleb Ziemis, Min-Yen Kan, Nancy F. Chen, Zhengyuan Liu and Diyi Yang

Optimizing Retrieval-augmented Reader Models via Token Elimination
Moshe Berchansky, Peter Izsak, Avi Caciularu, Ido Dagan and Moshe Wasserblat

WSDMS: Debunk Fake News via Weakly Supervised Detection of Misinforming Sentences with Contextualized Social Wisdom
Ruichao Yang, Wei Gao, Jing Ma, Hongzhan Lin and Zhiwei Yang

Robust Prompt Optimization for Large Language Models Against Distribution Shifts
Moxin Li, Wenjie Wang, Fuli Feng, Yixin Cao, Jizhi Zhang and Tat-Seng Chua

Exploiting Asymmetry for Synthetic Training Data Generation: SynthIE and the Case of Information Extraction
Martin Josifoski, Marija Sakota, Maxime Peyrard and Robert West
Condensing Multilingual Knowledge with Lightweight Language-Specific Modules
Haoran Xu, Weiting Tan, Shuyue Stella Li, Yunmo Chen, Benjamin Van Durme, Philipp Koehn and Kenton Murray ................................................................. 1575

The Framework Tax: Disparities Between Inference Efficiency in NLP Research and Deployment
Jared Fernandez, Jacob Kahn, Clara Na, Yonatan Bisk and Emma Strubell .......................... 1588

Evaluating Cross-Domain Text-to-SQL Models and Benchmarks
Mohammadreza Pourreza and Davood Rafiei ................................................................. 1601

Increasing Coverage and Precision of Textual Information in Multilingual Knowledge Graphs
Simone Conia, Min Li, Daniel Lee, Umar Farooq Minhas, Ihab Ilyas and Yunyao Li ....... 1612

Memory-Based Invariance Learning for Out-of-Domain Text Classification
Chen Jia and Yue Zhang ......................................................... 1635

Outlier Suppression+: Accurate quantization of large language models by equivalent and effective shifting and scaling
Xiuying Wei, Yunchen Zhang, Yuhang Li, Xiangguo Zhang, Ruihao Gong, Jinyang Guo and Xianglong Liu ................................................................. 1648

Three Stream Based Multi-level Event Contrastive Learning for Text-Video Event Extraction
Jiaqi Li, Chuanyi Zhang, Miaoze Eu, Dehai Min, Yongrui Chen and Guilin Qi ................. 1666

Diversify Question Generation with Retrieval-Augmented Style Transfer
Qi Gou, Zehua Xia, Bowen Yu, Haiyang Yu, Fei Huang, Yongbin Li and Nguyen Cam-Tu ... 1677

Fast and Accurate Factual Inconsistency Detection Over Long Documents
Barrett Martin Lattimer, Patrick Chen, Xinyuan Zhang and Yi Yang .............................. 1691

Interpreting Embedding Spaces by Conceptualization
Adi Simhi and Shaul Markovitch ................................................................................... 1704

Knowledge-Augmented Language Model Verification
Jinheon Baek, Soyeong Jeong, Minki Kang, Jong C. Park and Sung Ju Hwang ............... 1720

A Generation-based Deductive Method for Math Word Problems
Yuxuan Hu, Jing Zhang, Haoyang Li, Cuiping Li and Hong Chen ............................... 1737

Failures Pave the Way: Enhancing Large Language Models through Tuning-free Rule Accumulation
Zeyuan Yang, Peng Li and Yang Liu .............................................................................. 1751

Building Persona Consistent Dialogue Agents with Offline Reinforcement Learning
Ryan Shea and Zhou Yu .............................................................................................. 1778

Augmenting Zero-Shot Dense Retrievers with Plug-in Mixture-of-Memories
Suyu Ge, Chenyan Xiong, Corby Rosset, Arnold Overwijk, Jiawei Han and Paul N. Bennett 1796

Active Instruction Tuning: Improving Cross-Task Generalization by Training on Prompt Sensitive Tasks
Po-Nien Kung, Fan Yin, Di Wu, Kai-Wei Chang and Nanyun Peng ................................ 1813

Towards Example-Based NMT with Multi-Levenshtein Transformers
Maxime Bouthors, Josep Crego and François Yvon .................................................. 1830

DUnE: Dataset for Unified Editing
Afra Feyza Akyürek, Eric L Pan, Garry Kuwanto and Derry Wijaya ............................. 1847
CRT-QA: A Dataset of Complex Reasoning Question Answering over Tabular Data
Zhehao Zhang, Xitao Li, Yan Gao and Jian-Guang Lou ........................................... 2131

Promoting Topic Coherence and Inter-Document Consorts in Multi-Document Summarization via Simplicial Complex and Sheaf Graph
Yash Kumar Atri, Arun Iyer, Tanmoy Chakraborty and Vikram Goyal .......................... 2154

MAGNIFICo: Evaluating the In-Context Learning Ability of Large Language Models to Generalize to Novel Interpretations
Arkil Patel, Satwik Bhattamishra, Siva Reddy and Dzmitry Bahdanau ............................ 2167

Generating and Evaluating Tests for K-12 Students with Language Model Simulations: A Case Study on Sentence Reading Efficiency
Eric Zelikman, Wanjing Anya Ma, Jasmine Elizabeth Tran, Diyi Yang, Jason D Yeatman and Nick Haber .......................................................... 2190

Counter Turing Test (CT2): AI-Generated Text Detection is Not as Easy as You May Think - Introducing AI Detectability Index (ADI)
Megha Chakraborty, S.M Towhidul Islam Tonmoy, S M Mehedi Zaman, Shreya Gautam, Tanay Kumar, Krish Sharma, Niyar R Barman, Chandan Gupta, Vinija Jain, Aman Chadha, Amit P. Sheth and Amitava Das .......................................................... 2206

Revisiting the Optimality of Word Lengths
Tiago Pimentel, Clara Meister, Ethan Wilcox, Kyle Mahowald and Ryan Cotterell .......... 2240

Document-level Relationship Extraction by Bidirectional Constraints of Beta Rules
Yichun Liu, Zizhong Zhu, Xiaowang Zhang, Zhiyong Feng, Daoqi Chen and Yaxin Li ........ 2256

Instructed Language Models with Retrievers Are Powerful Entity Linkers
Zilin Xiao, Ming Gong, Jie Wu, Xingyao Zhang, Linjun Shou and Daxin Jiang ............... 2267

Towards Noise-Tolerant Speech-Referring Video Object Segmentation: Bridging Speech and Text
Xiang Li, Jinglu Wang, Xiaohao Xu, Muqiao Yang, Fan Yang, Yizhou Zhao, Rita Singh and Bhiksha Raj .......................................................... 2283

PROSE: A Pronoun Omission Solution for Chinese-English Spoken Language Translation
Ke Wang, Xiutian Zhao, Yanghui Li and Wei Peng ..................................................... 2297

A Diachronic Analysis of Paradigm Shifts in NLP Research: When, How, and Why?
Aniket Pramanick, Yufang Hou, Saif M. Mohammad and Iryna Gurevych ..................... 2312

Does the Correctness of Factual Knowledge Matter for Factual Knowledge-Enhanced Pre-trained Language Models?
Boxi Cao, Qiaoyu Tang, Hongyu Lin, Xianpei Han and Le Sun ........................................ 2327

Syntactic Substitutability as Unsupervised Dependency Syntax
Jasper Jian and Siva Reddy .................................................................................. 2341

MProto: Multi-Prototype Network with Denoised Optimal Transport for Distantly Supervised Named Entity Recognition
Shuhui Wu, Yongliang Shen, Zeqi Tan, Wenqi Ren, Jietian Guo, Shiliang Pu and Weiming Lu 2361

The Shifted and The Overlooked: A Task-oriented Investigation of User-GPT Interactions
Siru Ouyang, Shuohang Wang, Yang Liu, Ming Zhong, Yizhu Jiao, Dan Iter, Reid Pryzant, Chen-guang Zhu, Heng Ji and Jiawei Han .......................................................... 2375
Learning the Visualness of Text Using Large Vision-Language Models
Gaurav Verma, Ryan A. Rossi, Christopher Tensmeyer, Jiuxiang Gu and Ani Nenkova ........ 2394

The Past, Present and Better Future of Feedback Learning in Large Language Models for Subjective Human Preferences and Values
Hannah Rose Kirk, Andrew Michael Bean, Berrie Vidgen, Paul Rottger and Scott A. Hale .... 2409

TempTabQA: Temporal Question Answering for Semi-Structured Tables
Vivek Gupta, Pranshu Kandoi, Mahesh Bhavesh Vora, Shuo Zhang, Yujie He, Ridho Reinanda and Vivek Srikumar ............................................................... 2431

Task-Level Thinking Steps Help Large Language Models for Challenging Classification Task
Chunhui Du, Jidong Tian, Haoran Liao, Jindou Chen, Hao He and Yaohui Jin ............... 2454

RepoCoder: Repository-Level Code Completion Through Iterative Retrieval and Generation
Fengji Zhang, Bei Chen, Yue Zhang, Jacky Keung, Jin Liu, Daoguang Zan, Yi Mao, Jian-Guang Lou and Weizhu Chen ................................................................. 2471

Influence Scores at Scale for Efficient Language Data Sampling
Nikhil Anand, Joshua Tan and Maria Minakova ...................................................... 2485

G-Eval: NLG Evaluation using Gpt-4 with Better Human Alignment
Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu and Chenguang Zhu ...... 2511

Learning Retrieval Augmentation for Personalized Dialogue Generation
Qiushi Huang, Shuai Fu, Xubo Liu, Wenhao Wang, Tom Ko, Yu Zhang and Lilian Tang .... 2523

The Troubling Emergence of Hallucination in Large Language Models - An Extensive Definition, Quantification, and Prescriptive Remediations
Vipula Rawte, Swagata Chakraborty, Agnihot Pathak, Anubhav Sarkar, S.M Towhidul Islam Tonomoy, Aman Chadha, Amit P. Sheth and Amitava Das .............................................. 2541

NAIL: Lexical Retrieval Indices with Efficient Non-Autoregressive Decoders
Livio Baldini Soares, Daniel Gillick, Jeremy R. Cole and Tom Kwiatkowski ............... 2574

Analyzing Modular Approaches for Visual Question Decomposition
Apoorv Khandelwal, Ellie Pavlick and Chen Sun .................................................. 2590

Improving Summarization with Human Edits
Zonghai Yao, Benjamin J Schloss and Sai P Selvaraj ............................................. 2604

Did You Mean...? Confidence-based Trade-offs in Semantic Parsing
Elias Stengel-Eskin and Benjamin Van Durme ..................................................... 2621

The Skipped Beat: A Study of Sociopragmatic Understanding in LLMs for 64 Languages
Chiyu Zhang, Khai Duy Doan, Qisheng Liao and Muhammad Abdul-Mageed .......... 2630

Understanding the Effect of Model Compression on Social Bias in Large Language Models
Gustavo Gonçalves and Emma Strubell ................................................................. 2663

BioPlanner: Automatic Evaluation of LLMs on Protocol Planning in Biology
Odhran O’Donoghue, Aleksandar Shtedritski, John Ginger, Ralph Abboud, Ali Essam Ghaereeb and Samuel G Rodrigues ......................................................... 2676

Cross-lingual Prompting: Improving Zero-shot Chain-of-Thought Reasoning across Languages
Libo Qin, Qiguang Chen, Fuxuan Wei, Shijue Huang and Wanxiang Che .................. 2695
FinGPT: Large Generative Models for a Small Language

Boosting Summarization with Normalizing Flows and Aggressive Training
Yu Yang and Xiaotong Shen ........................................... 2727

Indicative Summarization of Long Discussions
Shahbaz Syed, Dominik Schwabe, Khalid Al Khatib and Martin Potthast .............. 2752

A Framework for Vision-Language Warm-up Tasks in Multimodal Dialogue Models
Jaewook Lee, Seongsik Park, Seong-Heum Park, Hongjin Kim and Harksoo Kim ........ 2789

Once is Enough: A Light-Weight Cross-Attention for Fast Sentence Pair Modeling
Yuanhang Yang, Shiyi Qi, Chuanyi Liu, Qifan Wang, Cuiyun Gao and Zenglin Xu ........ 2800

Plan, Verify and Switch: Integrated Reasoning with Diverse X-of-Thoughts
Tengxiao Liu, Qipeng Guo, Yuqing Yang, Xiangkun Hu, Yue Zhang, Xipeng Qiu and Zheng Zhang .......................................................... 2807

GLEN: General-Purpose Event Detection for Thousands of Types
Sha Li, Qiusi Zhan, Kathryn Conger, Martha Palmer, Heng Ji and Jiawei Han ........... 2823

Hierarchical Pretraining on Multimodal Electronic Health Records
Xiaochen Wang, Junyu Luo, Jiaqi Wang, Ziyi Yin, Suhan Cui, Yuan Zhong, Yaqing Wang and Fenglong Ma ...................................................... 2839

Critic-Driven Decoding for Mitigating Hallucinations in Data-to-text Generation
Mateusz Lango and Ondrej Dusek ........................................ 2853

Bridging the Gap between Synthetic and Authentic Images for Multimodal Machine Translation
Wenyu Guo, Qingkai Fang, Dong Yu and Yang Feng .................................. 2863

DEPN: Detecting and Editing Privacy Neurons in Pretrained Language Models
Xinwei Wu, Junzhuo Li, Minghui Xu, Weilong Dong, Shuangzhi Wu, Chao Bian and Deyi Xiong 2875

Investigating Bias in Multilingual Language Models: Cross-Lingual Transfer of Debiasing Techniques
Manon Reusens, Philipp Borchert, Margot Mieskes, Jochen De Weerdt and Bart Baesens . 2887

Can Language Models Laugh at YouTube Short-form Videos?
Dayoon Ko, Sangho Lee and Gunhee Kim ....................................... 2897

Random Entity Quantization for Parameter-Efficient Compositional Knowledge Graph Representation
Jiaang Li, Quan Wang, Yi Liu, Licheng Zhang and Zhendong Mao .......................... 2917

Exploring All-In-One Knowledge Distillation Framework for Neural Machine Translation
Zhongjian Miao, Wen Zhang, Jinsong Su, Xiang Li, Jian Luan, Yidong Chen, Bin Wang and Min Zhang ......................................................... 2929

HistAlign: Improving Context Dependency in Language Generation by Aligning with History
David Wan, Shiyue Zhang and Mohit Bansal ....................................... 2941

CombLM: Adapting Black-Box Language Models through Small Fine-Tuned Models
Aitor Ormazabal, Mikel Artetxe and Eneko Agirre ................................. 2961

xlsvii
Enhancing Low-resource Fine-grained Named Entity Recognition by Leveraging Coarse-grained Datasets
Su Ah Lee, Seokjin Oh and Woohwan Jung .................................................. 3269

Oolong: Investigating What Makes Transfer Learning Hard with Controlled Studies
Zhengxuan Wu, Alex Tamkin and Isabel Papadimitriou .................................. 3280

Non-Autoregressive Math Word Problem Solver with Unified Tree Structure
Yi Bin, Mengqun Han, Wenhao Shi, Lei Wang, Yang Yang, See-Kiong Ng and Heng Tao Shen 3290

Improving Chinese Pop Song and Hokkien Gezi Opera Singing Voice Synthesis by Enhancing Local Modeling
Peng Bai, Yue Zhou, Meizhen Zheng, Wujin Sun and Xiaodong Shi .......................... 3302

What Else Do I Need to Know? The Effect of Background Information on Users’ Reliance on QA Systems
Navita Goyal, Eleftheria Briakou, Amanda Stephanie Liu, Connor Baumler, Claire Bonial, Jeffrey Micher, Clare R. Voss, Marine Carpuat and Hal Daumé III ................................. 3313

GROOViST: A Metric for Grounding Objects in Visual Storytelling
Aditya Kaushik Surikuchi, Sandro Pezzelle and Raquel Fernández ...................... 3331

VIBE: Topic-Driven Temporal Adaptation for Twitter Classification
Yuji Zhang, Jing Li and Wenjie Li ................................................................. 3340

TOD-Flow: Modeling the Structure of Task-Oriented Dialogues
Sungryull Sohn, Yiwei Lyu, Anthony Zhe Liu, Lajanugen Logeswaran, Dong-Ki Kim, Dongsun Shim and Honglak Lee ............................................................... 3355

TopWORDS-Poetry: Simultaneous Text Segmentation and Word Discovery for Classical Chinese Poetry via Bayesian Inference
Changzai Pan, Feiyue Li and Ke Deng .......................................................... 3372

Knowledge Rumination for Pre-trained Language Models
Yunzhi Yao, Peng Wang, Shengyu Mao, Chuanqi Tan, Fei Huang, Huajun Chen and Ningyu Zhang ........................................................................................................... 3387

Struct-XLM: A Structure Discovery Multilingual Language Model for Enhancing Cross-lingual Transfer through Reinforcement Learning
Linjuan Wu and Weiming Lu ........................................................................ 3405

AdaSent: Efficient Domain-Adapted Sentence Embeddings for Few-Shot Classification
Yongxin Huang, Kexin Wang, Sourav Dutta, Raj Nath Patel, Goran Glavaš and Iryna Gurevych 3420

Interview Evaluation: A Novel Approach for Automatic Evaluation of Conversational Question Answering Models
Xibo Li, Bowei Zou, Yifan Fan, Yanling Li, Ai Ti Aw and Yu Hong ...................... 3435

TCFLE-8: a Corpus of Learner Written Productions for French as a Foreign Language and its Application to Automated Essay Scoring
Rodrigo Wilkens, Alice Pintard, David Alfter, Vincent Folny and Thomas François .......... 3447

Dancing Between Success and Failure: Edit-level Simplification Evaluation using SALSA
David Heineman, Yao Dou, Mounica Maddela and Wei Xu .............................. 3466
Confidence-based Ensembling of Perspective-aware Models
Silvia Casola, Soda Marem Lo, Valerio Basile, Simona Frenda, Alessandra Teresa Cignarella, Viviana Patti and Cristina Bosco

ToViLaG: Your Visual-Language Generative Model is Also An Evildoer
Xinpeng Wang, Xiaoyuan Yi, Han Jiang, Shanlin Zhou, Zhihua Wei and Xing Xie

GPT-RE: In-context Learning for Relation Extraction using Large Language Models
Zhen Wan, Fei Cheng, Zhuoyuan Mao, Qianying Liu, Haiyue Song, Jiwei Li and Sadao Kurohashi

Sociocultural Norm Similarities and Differences via Situational Alignment and Explainable Textual Entailment
Sky CH-Wang, Arkadiy Saakyan, Oliver Li, Zhou Yu and Smaranda Muresan

INFORM: Information Entropy based multi-step reasoning FOR large language Models
Chuyue Zhou, Wangjie You, Juntao Li, Jing Ye, Kehai Chen and Min Zhang

Adaptive Gating in Mixture-of-Experts based Language Models
Jiamin Li, Qiang Su, Yitao Yang, Yimin Jiang, Cong Wang and Hong Xu

On the Automatic Generation and Simplification of Children’s Stories
Maria Valentini, Jennifer Weber, Jesus Salcido, Téa Wright, Eliana Colunga and Katharina von der Wense

When Do Decompositions Help for Machine Reading?
Kangda Wei, Dawn Lawrie, Benjamin Van Durme, Yunmo Chen and Orion Weller

The Curious Case of Hallucinatory (Un)answerability: Finding Truths in the Hidden States of Over-Confident Large Language Models
Aviv Slobodkin, Omer Goldman, Avi Caciularu, Ido Dagan and Shauli Ravfogel

Identifying Informational Sources in News Articles
Alexander Spangher, Nanyun Peng, Emilio Ferrara and Jonathan May

Retrofitting Light-weight Language Models for Emotions using Supervised Contrastive Learning
Sapan Shah, Sreedhar Reddy and Pushpak Bhattacharyya

Longtriever: a Pre-trained Long Text Encoder for Dense Document Retrieval
Junhan Yang, Zheng Liu, Chaozhuo Li, Guangzhong Sun and Xing Xie

Revisiting De-Identification of Electronic Medical Records: Evaluation of Within- and Cross-Hospital Generalization
Yiyang Liu, Jinpeng Li and Enwei Zhu

Small Language Models Fine-tuned to Coordinate Larger Language Models improve Complex Reasoning
Gurusha Juneja, Subhabrata Dutta, Soumen Chakrabarti, Sunny Manchanda and Tam moy Chakraborty

Language Representation Projection: Can We Transfer Factual Knowledge across Languages in Multilingual Language Models?
Shaoyang Xu, Junzhuo Li and Deyi Xiong

Structural Priming Demonstrates Abstract Grammatical Representations in Multilingual Language Models
James Michaelov, Catherine Arnett, Tyler A. Chang and Ben Bergen
ReasoningLM: Enabling Structural Subgraph Reasoning in Pre-trained Language Models for Question Answering over Knowledge Graph
Jinhao Jiang, Kun Zhou, Xin Zhao, Yaliang Li and Ji-Rong Wen .............................................. 3721

Deep Natural Language Feature Learning for Interpretable Prediction
Felipe Urrutia, Cristian Buc Calderon and Valentin Barriera .................................................. 3736

ROBBIE: Robust Bias Evaluation of Large Generative Language Models
David Esiobu, Xiaoping Tan, Saghar Hosseini, Megan Ung, Yuchen Zhang, Jude Fernandes, Jane Dwivedi-Yu, Eleonora Presani, Adina Williams and Eric Michael Smith ............................................ 3764

Enhancing Task-oriented Dialogue Systems with Generative Post-processing Networks
Atsumoto Ohashi and Ryuichiro Higashinaka ................................................................. 3815

Adapting Language Models to Compress Contexts
Alexis Chevalier, Alexander Wettig, Anirudh Ajith and Danqi Chen ........................................ 3829

Yichao Zhou, James Bradley Wendt, Navneet Potti, Jing Xie and Sandeep Tata ..................... 3847

TRAVEL: Tag-Aware Conversational FAQ Retrieval via Reinforcement Learning
Yue Chen, Dingnan Jin, Chen Huang, Jia Liu and Wenqiang Lei .............................................. 3861

Continual Dialogue State Tracking via Example-Guided Question Answering
Hyundong Justin Cho, Andrea Madotto, Zhaojiang Lin, Khyathi Chandu, Satwik Kottur, Jing Xu, Jonathan May and Chinmudurui Sankar ............................................................. 3873

Lost in Translation, Found in Spans: Identifying Claims in Multilingual Social Media
Shubham Mittal, Megha Sundriyal and Preslav Nakov ...................................................... 3887

COVID-19 Vaccine Misinformation in Middle Income Countries
Jongin Kim, Byeod Rhee Bak, Aditya Agrawal, Jiaxi Wu, Veronika J. Wirtz, Traci Hong and Derry Wijaya .......................................................... 3903

Contrastive Learning of Sentence Embeddings from Scratch
Junlei Zhang, Zhenzhong Lan and Junxian He ........................................................................ 3916

A Rose by Any Other Name would not Smell as Sweet: Social Bias in Names Mistranslation
Sandra Camille Sandoval, Jieyu Zhao, Marine Carpuat and Hal Daumé III .......................... 3933

Investigating Efficiently Extending Transformers for Long Input Summarization
Jason Phang, Yao Zhao and Peter J Liu ............................................................................. 3946

CS2W: A Chinese Spoken-to-Written Style Conversion Dataset with Multiple Conversion Types
Zishan Guo, Linhao Yu, Minghui Xu, Renren Jin and Deyi Xiong ....................................... 3962

Unifying Cross-Lingual Transfer across Scenarios of Resource Scarcity
Alan Ansell, Marinela Parović, Ivan Vulić, Anna Korhonen and Edoardo Ponti ................. 3980

A Tale of Pronouns: Interpretability Informs Gender Bias Mitigation for Fairer Instruction-Tuned Machine Translation
Giuseppe Attanasio, Flor Miriam Plaza del Arco, Debora Nozza and Anne Lauscher .......... 3996

DisCo: Distilled Student Models Co-training for Semi-supervised Text Mining
Weifeng Jiang, Qianren Mao, Chenghua Lin, Jianxin Li, Ting Deng, Weiyi Yang and Zheng Wang 4015
Uncertainty Guided Global Memory Improves Multi-Hop Question Answering
Alsu Sagirova and Mikhail Burtsev ............................... 4317

Prompting Large Language Models with Chain-of-Thought for Few-Shot Knowledge Base Question Generation
Yuanyuan Liang, Jianing Wang, Hanlun Zhu, Lei Wang, Weining Qian and Yunshi Lan .... 4329

TrojanSQL: SQL Injection against Natural Language Interface to Database
Jinchuan Zhang, Yan Zhou, Binyuan Hui, Yaxin Liu, Ziming Li and Songlin Hu ............ 4344

Preserving Privacy Through Dememorization: An Unlearning Technique For Mitigating Memorization Risks In Language Models
Aly M. Kassem, Omar Mahmoud and Sherif Saad ................................. 4360

MingOfficial: A Ming Official Career Dataset and a Historical Context-Aware Representation Learning Framework
You-Jun Chen, Hsin-Yi Hsieh, Yu Tung Lin, Yingtao Tian, Bert Chan, Yu-Sin Liu, Yi-Hsuan Lin and Richard Tzong-Han Tsai ................................. 4380

DPP-TTS: Diversifying prosodic features of speech via determinantal point processes
Seongho Joo, Hyukhun Koh and Kyomin Jung .................................. 4402

Meta-Learning Online Adaptation of Language Models
Nathan Zixia Hu, Eric Mitchell, Christopher D Manning and Chelsea Finn ................................. 4418

Self-Detoxifying Language Models via Toxification Reversal
Chak Tou Leong, Yi Cheng, Jiashuo Wang, Jian Wang and Wenjie Li ................................. 4433

Interactive Text Generation
Felix Faltings, Michel Galley, Kianté Brantley, Baolin Peng, Weixin Cai, Yizhe Zhang, Jianfeng Gao and Bill Dolan .................................. 4450

Knowledge Distillation ≈ Label Smoothing: Fact or Fallacy?
Md Arafat Sultan .......................................................... 4469

Analyzing Cognitive Plausibility of Subword Tokenization
Lisa Beinborn and Yuval Pinter ........................................ 4478

POE: Process of Elimination for Multiple Choice Reasoning
Chenkai Ma and Xinya Du .................................................. 4487

NeuSTIP: A Neuro-Symbolic Model for Link and Time Prediction in Temporal Knowledge Graphs
Ishaan Singh, Navdeep Kaur, Garima Gaur and Mausam ........................................ 4497

Standardizing Distress Analysis: Emotion-Driven Distress Identification and Cause Extraction (DICE) in Multimodal Online Posts
Gopendra Vikram Singh, Soumitra Ghosh, Atul Verma, Chetna Painkra and Asif Ekbal ...... 4517

Out-of-Distribution Generalization in Natural Language Processing: Past, Present, and Future
Linyi Yang, Yaoxian Song, Xuan Ren, Chenyang Lyu, Yidong Wang, Jingming Zhuo, Lingqiao Liu, Jindong Wang, Jennifer Foster and Yue Zhang ........................................ 4533

Noisy Exemplars Make Large Language Models More Robust: A Domain-Agnostic Behavioral Analysis
Hongyi Zheng and Abulhair Saparov ........................................ 4560

Can Large Language Models Capture Dissenting Human Voices?
Noah Lee, Na Min An and James Thorne ........................................ 4569
DecoMT: Decomposed Prompting for Machine Translation Between Related Languages using Large Language Models
Ratish Puduppully, Anoop Kunchukuttan, Raj Dabre, Ai Ti Aw and Nancy F. Chen 4586

Prototype-based HyperAdapter for Sample-Efficient Multi-task Tuning
Hao Zhao, Jie Fu and Zhaofeng He 4603

Towards Building More Robust NER datasets: An Empirical Study on NER Dataset Bias from a Dataset Difficulty View
Ruotian Ma, Xiaolei Wang, Xin Zhou, Qi Zhang and Xuanjing Huang 4616

GradSim: Gradient-Based Language Grouping for Effective Multilingual Training
Mingyang Wang, Heike Adel, Lukas Lange, Jannik Strötgen and Hinrich Schuetze 4631

Discovering Universal Geometry in Embeddings with ICA
Hiroaki Yamagiwa, Momose Oyama and Hidetoshi Shimodaira 4647

Toward a Critical Toponymy Framework for Named Entity Recognition: A Case Study of Airbnb in New York City
Mikael Brunila, Jack LaViolette, Sky CH-Wang, Priyanka Verma, Clara Fére and Grant McKenzie 4676

Well Begun is Half Done: Generator-agnostic Knowledge Pre-Selection for Knowledge-Grounded Dialogue
Lang Qin, Yao Zhang, Hongru Liang, Jun Wang and Zhenglu Yang 4696

Merging Generated and Retrieved Knowledge for Open-Domain QA
Yunxiang Zhang, Muhammad Khalifa, Lajanugen Logeswaran, Moontae Lee, Honglak Lee and Lu Wang 4710

Best of Both Worlds: Towards Improving Temporal Knowledge Base Question Answering via Targeted Fact Extraction
Nithish Kannen, Udit Sharma, Sumit Neelam, Dinesh Khandelwal, Shajith Ikbal, Hima Karanam and L Venkata Subramaniam 4729

Text Fact Transfer
Nishant Balepur, Jie Huang and Kevin Chang 4745

A Cheaper and Better Diffusion Language Model with Soft-Masked Noise
Jiaao Chen, Aston Zhang, Mu Li, Alex Smola and Diyi Yang 4765

Mirages. On Anthropomorphism in Dialogue Systems
Gavin Abercrombie, Amanda Cercas Curry, Tanvi Dinkar, Verena Rieser and Zeerak Talat 4776

Cognitive Dissonance: Why Do Language Model Outputs Disagree with Internal Representations of Truthfulness?
Kevin Liu, Stephen Casper, Dylan Hadfield-Menell and Jacob Andreas 4791

KEBAP: Korean Error Explainable Benchmark Dataset for ASR and Post-processing
Seonmin Koo, Chanjun Park, Jinsung Kim, Jaehyung Seo, Suygeon Eo, Hyeonseok Moon and Heuiseok Lim 4798

Adaptive Policy with Wait-k Model for Simultaneous Translation
Libo Zhao, Kai Fan, Wei Luo, Wu Jing, Shushu Wang, Ziqian Zeng and Zhongqiang Huang 4816

Cross-Document Event Coreference Resolution on Discourse Structure
Xinyu Chen, Sheng Xu, Peifeng Li and Qiaoming Zhu 4833
Post-hoc Utterance Refining Method by Entity Mining for Faithful Knowledge Grounded Conversations
Yoonna Jang, Suhyune Son, Jeongwoo Lee, Junyoung Son, Yuna Hur, Jungwoo Lim, Hyeonseok Moon, Kisu Yang and Heuiseok Lim ....................................................... 4844

Can We Edit Factual Knowledge by In-Context Learning?
Ce Zheng, Lei Li, Qingxiu Dong, Yuxuan Fan, Zhiyong Wu, Jingjing Xu and Baobao Chang 4862

EDIS: Entity-Driven Image Search over Multimodal Web Content
Siqi Liu, Weixi Feng, Tsu-Jui Fu, Wenhu Chen and William Yang Wang ....................... 4877

GQA: Training Generalized Multi-Query Transformer Models from Multi-Head Checkpoints
Joshua Ainslie, James Lee-Thorp, Michiel de Jong, Yury Zemlyanskiy, Federico Lebron and Sumit Sanghai ................................................................. 4895

Towards a Mechanistic Interpretation of Multi-Step Reasoning Capabilities of Language Models
Yifan Hou, Jiaoda Li, Yu Fei, Alessandro Stolfo, Wangchunshu Zhou, Guangtao Zeng, Antoine Bosselut and Mrinmaya Sachan ........................................... 4902

BiasX: “Thinking Slow” in Toxic Content Moderation with Explanations of Implied Social Biases
Yiming Zhang, Sravani Uttara Nanduri, Liwei Jiang, Tongshuang Wu and Maarten Sap .... 4920

Text encoders bottleneck compositionality in contrastive vision-language models
Amita Kamath, Jack Hessel and Kai-Wei Chang .......................................................... 4933

Ignore This Title and HackAPrompt: Exposing Systemic Vulnerabilities of LLMs Through a Global Prompt Hacking Competition
Sander V Schulhoff, Jeremy Pinto, Anaum Khan, Louis-François Bouchard, Chenglei Si, Svetlina Anati, Valen Tagliabue, Anson Liu Kost, Christopher R Carnahan and Jordan Lee Boyd-Graber .. 4945

MMNMT: Modularizing Multilingual Neural Machine Translation with Flexibly Assembled MoE and Dense Blocks
Shangjie Li, Xiangpeng Wei, Shaolin Zhu, Jun Xie, Baosong Yang and Deyi Xiong ......... 4978

Localizing Active Objects from Egocentric Vision with Symbolic World Knowledge
Te-Lin Wu, Yu Zhou and Nanyun Peng ..................................................................... 4991

Introducing Rhetorical Parallelism Detection: A New Task with Datasets, Metrics, and Baselines
Stephen Bothwell, Justin DeBenedetto, Theresa Crnkovich, Hildegund Muller and David Chiang 5007

Prompting is not a substitute for probability measurements in large language models
Jennifer Hu and Roger P. Levy ................................................................. 5040

Parameter-Efficient Language Model Tuning with Active Learning in Low-Resource Settings
Josip Jukić and Jan Snojder ............................................................. 5061

Stop Uploading Test Data in Plain Text: Practical Strategies for Mitigating Data Contamination by Evaluation Benchmarks
Alon Jacovi, Avi Caciularu, Omer Goldman and Yoav Goldberg ......................... 5075

CoLT5: Faster Long-Range Transformers with Conditional Computation
Joshua Ainslie, Tao Lei, Michiel de Jong, Santiago Ontanon, Siddhartha Brahma, Yury Zemlyanskiy, David Uthus, Mandy Guo, James Lee-Thorp, Yi Tay, Yun-Hsuan Sung and Sumit Sanghai 5085

DiSTRICT: Dialogue State Tracking with Retriever Driven In-Context Tuning
Praveen Venkateswaran, Evelyn Duesterwald and Vatche Isahagian ....................... 5101
Cross-Cultural Analysis of Human Values, Morals, and Biases in Folk Tales
Winston Wu, Lu Wang and Rada Mihalcea .................................................. 5113

Non-Programmers Can Label Programs Indirectly via Active Examples: A Case Study with Text-to-SQL
Ruiqi Zhong, Charlie Victor Snell, Dan Klein and Jason Eisner ......................... 5126

LINC: A Neurosymbolic Approach for Logical Reasoning by Combining Language Models with First-Order Logic Provers
Theo X. Olausson, Alex Gu, Ben Lipkin, Cedegao E. Zhang, Armando Solar-Lezama, Joshua B. Tenenbaum and Roger P. Levy .................................................. 5153

Non-autoregressive Streaming Transformer for Simultaneous Translation
Zhengrui Ma, Shaolei Zhang, Shoutao Guo, Chenze Shao, Min Zhang and Yang Feng ...... 5177

ViSoBERT: A Pre-Trained Language Model for Vietnamese Social Media Text Processing
Nam Quoc Nguyen, Thang Chau Phan, Duc-Vu Nguyen and Kiet Van Nguyen ........... 5191

RAPL: A Relation-Aware Prototype Learning Approach for Few-Shot Document-Level Relation Extraction
Shiao Meng, Xuming Hu, Aiwei Liu, Shuang Li, Fukun Ma, Yawen Yang and Lijie Wen .... 5208

GeoLM: Empowering Language Models for Geospatially Grounded Language Understanding
Zekun Li, Wenxuan Zhou, Yao-Yi Chiang and Muhao Chen .................................. 5227

Cross-Modal Conceptualization in Bottleneck Models
Danis Alukaev, Semen Kiselev, Ilya Pershin, Bulat Ibragimov, Vladimir V. Ivanov, Alexey Kornaev and Ivan Titov .......................................................... 5241

LLM-Adapters: An Adapter Family for Parameter-Efficient Fine-Tuning of Large Language Models
Zhiqiang Hu, Lei Wang, Yihuai Lan, Wanyu Xu, Ee-Peng Lim, Lidong Bing, Xing Xu, Soujanya Poria and Roy Ka-Wei Lee .............................................. 5254

DREAM: Deployment of Recombination and Ensembles in Argument Mining
Florian Ruosch, Cristina Sarasua and Abraham Bernstein .................................. 5277

MILDSum: A Novel Benchmark Dataset for Multilingual Summarization of Indian Legal Case Judgments
Debtanu Datta, Shubham Soni, Rajdeep Mukherjee and Saptarshi Ghosh .................... 5291

Query Rewriting in Retrieval-Augmented Large Language Models
Xinbei Ma, Yeyun Gong, Pengcheng He, Hai Zhao and Nan Duan ............................ 5303

PromptMix: A Class Boundary Augmentation Method for Large Language Model Distillation
Gaurav Sahu, Olga Vechtomova, Dzmitry Bahdanau and Issam H. Laradji .................... 5316

COHESENTIA: A Novel Benchmark of Incremental versus Holistic Assessment of Coherence in Generated Texts
Aviya Maimon and Reut Tsarfaty ................................................................. 5328

QUDeval: The Evaluation of Questions Under Discussion Discourse Parsing
Yating Wu, Ritika Rajesh Mangla, Greg Durrett and Junyi Jessy Li ............................ 5344

PRCA: Fitting Black-Box Large Language Models for Retrieval Question Answering via Pluggable Reward-Driven Contextual Adapter
Haoyan Yang, Zhitao Li, Yong Zhang, Jianzong Wang, Ning Cheng, Ming Li and Jing Xiao 5364

Exploring Chain of Thought Style Prompting for Text-to-SQL
Chang-Yu Tai, Ziru Chen, Tianshu Zhang, Xiang Deng and Huan Sun ...................... 5376
Efficient Algorithms for Recognizing Weighted Tree-Adjoining Languages
Alexandra Butoi, Tim Vieira, Ryan Cotterell and David Chiang 5394

Harnessing Black-Box Control to Boost Commonsense in LM’s Generation
Yufei Tian, Felix Zhang and Nanyun Peng 5417

Just Ask for Calibration: Strategies for Eliciting Calibrated Confidence Scores from Language Models Fine-Tuned with Human Feedback
Katherine Tian, Eric Mitchell, Allan Zhou, Archit Sharma, Rafael Rafailov, Huaxiu Yao, Chelsea Finn and Christopher D Manning 5433

Representative Demonstration Selection for In-Context Learning with Two-Stage Determinantal Point Process
Zhao Yang, Yuanzhe Zhang, Dianbo Sui, Cao Liu, Jun Zhao and Kang Liu 5443

The Effect of Scaling, Retrieval Augmentation and Form on the Factual Consistency of Language Models
Lovisa Hagström, Denitsa Saynova, Tobias Norlund, Moa Johansson and Richard Johansson 5457

ViPE: Visualise Pretty-much Everything
Hassan Shahmohammadi, Adhiraj Ghosh and Hendrik Lensch 5477

Semi-automatic Data Enhancement for Document-Level Relation Extraction with Distant Supervision from Large Language Models
Junpeng Li, Zixia Jia and Zilong Zheng 5495

Navigating the Grey Area: How Expressions of Uncertainty and Overconfidence Affect Language Models
Kaitlyn Zhou, Dan Jurafsky and Tatsunori Hashimoto 5506

Elaborative Simplification as Implicit Questions Under Discussion
Yating Wu, William Berkeley Sheffield, Kyle Mahowald and Junyi Jessy Li 5525

EntSUMv2: Dataset, Models and Evaluation for More Abstractive Entity-Centric Summarization
Dhruv Mehra, Lingjue Xie, Ella Hofmann-Coyle, Mayank Kulkarni and Daniel Preotiuc-Pietro 5538

Amanpreet Singh, Mike D’Arcy, Arman Cohan, Doug Downey and Sergey Feldman 5548

A Diachronic Perspective on User Trust in AI under Uncertainty
Shehzaad Zuzar Dhuliawala, Vilém Zouhar, Mennatallah El-Assady and Mrinmaya Sachan 5567

CT-GAT: Cross-Task Generative Adversarial Attack based on Transferability
Minxuan Lv, Chengwei Dai, Kun Li, Wei Zhou and Songlin Hu 5581

Improve Long Document Topic Segmentation Models With Enhanced Coherence Modeling
Hai Yu, Chong Deng, Qinglin Zhang, Jiaqing Liu, Qian Chen and Wen Wang 5592

Dialogue Chain-of-Thought Distillation for Commonsense-aware Conversational Agents
Hyungjoo Chae, Yongho Song, Kai Tzu-iunn Ong, Taeyoon Kwon, Minjin Kim, Youngjae Yu, Dongha Lee, Dongyeop Kang and Jinyoung Yeo 5606

Information Value: Measuring Utterance Predictability as Distance from Plausible Alternatives
Mario Giulianelli, Sarenne Wallbridge and Raquel Fernández 5633

Generating Commonsense Counterfactuals for Stable Relation Extraction
Xin Miao, Yongqi Li and Tieyun Qian 5654
C-STS: Conditional Semantic Textual Similarity

Cross-lingual Transfer Can Worsen Bias in Sentiment Analysis
Seraphina Goldfarb-Tarrant, Björn Ross and Adam Lopez .......................... 5691

Rumor Detection on Social Media with Crowd Intelligence and ChatGPT-Assisted Networks
Chang Yang, Peng Zhang, Wenbo Qiao, Hui Gao and Jiaming Zhao ......................... 5705

Grounding Visual Illusions in Language: Do Vision-Language Models Perceive Illusions Like Humans?
Yichi Zhang, Jiayi Pan, Yuchen Zhou, Rui Pan and Joyce Chai .......................... 5718

Analysing State-Backed Propaganda Websites: a New Dataset and Linguistic Study
Freddy Heppell, Kalina Bontcheva and Carolina Scarton ................................ 5729

Controllable Contrastive Generation for Multilingual Biomedical Entity Linking
Tiantian Zhu, Yang Qin, Qingcai Chen, Xin Mu, Changlong Yu and Yang Xiang .......... 5742

HyperRouter: Towards Efficient Training and Inference of Sparse Mixture of Experts
Truong Giang Do, Le Huy Khiem, Quang Pham, Trung Tin Nguyen, Thanh Nam Doan, Binh T. Nguyen, Chenghao Liu, Savitha Ramasamy, Xiaoli Li and Steven Hoi .................................. 5754

MediaHG: Rethinking Eye-catching Features in Social Media Headline Generation
Boning Zhang and Yang Yang ........................................... 5766

Fine-tuned LLMs Know More, Hallucinate Less with Few-Shot Sequence-to-Sequence Semantic Parsing over Wikidata
Silei Xu, Shicheng Liu, Theo Culhane, Elizaveta Pertseva, Meng Hsi Wu, Sina Semnani and Monica Lam ................................................................. 5778

ZEROTOP: Zero-Shot Task-Oriented Semantic Parsing using Large Language Models
Dheeraj Mekala, Jason Andrew Wolfe and Subhro Roy .................................. 5792

Efficient Grammatical Error Correction Via Multi-Task Training and Optimized Training Schedule
Andrey Bout, Alexander Podolskiy, Sergey Nikolenko and Irina Piontkovskaya .......... 5800

The BLA Benchmark: Investigating Basic Language Abilities of Pre-Trained Multimodal Models
Xinyi Chen, Raquel Fernández and Sandro Pezzelle ........................................ 5817

RainProof: An Umbrella to Shield Text Generator from Out-Of-Distribution Data
Maxime Darrin, Pablo Piantanida and Pierre Colombo .................................. 5831

KEPL: Knowledge Enhanced Prompt Learning for Chinese Hypernym-Hyponym Extraction
Ningchen Ma, Dong Wang, Hongyun Bao, Lei He and Suncong Zheng ................. 5858

Ditto: A Simple and Efficient Approach to Improve Sentence Embeddings
Qian Chen, Wen Wang, Qinglin Zhang, Siqi Zheng, Chong Deng, Hai Yu, Jiaqing Liu, Yukun Ma and Chong Zhang ......................................................... 5868

Preserving Knowledge Invariance: Rethinking Robustness Evaluation of Open Information Extraction
Ji Qi, Chuchun Zhang, Xiaozhi Wang, Kaisheng Zeng, Jifan Yu, Jinxin Liu, Lei Hou, Jianzi Li and Xu Bin ................................................................. 5876

Why Should This Article Be Deleted? Transparent Stance Detection in Multilingual Wikipedia Editor Discussions
Lucie Aimée Kaffee, Arnav Arora and Isabelle Augenstein .............................. 5891
Fast and Robust Early-Exiting Framework for Autoregressive Language Models with Synchronized Parallel Decoding
Sangmin Bae, Jongwoo Ko, Hwanjun Song and Se-Young Yun .................................................. 5910

End-to-end Task-oriented Dialogue: A Survey of Tasks, Methods, and Future Directions
Libo Qin, Wenbo Pan, Qiguang Chen, Lizi Liao, Zhou Yu, Yue Zhang, Wanxiang Che and Min Li ................................................................. 5925

Answering Questions by Meta-Reasoning over Multiple Chains of Thought
Ori Yoran, Tomer Wolfson, Ben Bogin, Uri Katz, Daniel Deutch and Jonathan Berant ........... 5942

INSTRUCTSCORE: Towards Explainable Text Generation Evaluation with Automatic Feedback
Wenda Xu, Danqing Wang, Liangming Pan, Zhenqiao Song, Markus Freitag, William Yang Wang and Lei Li ................................................................. 5967

Multi-level Contrastive Learning for Script-based Character Understanding
Dawei Li, Hengyuan Zhang, Yanran Li and Shiping Yang ..................................................... 5995

CHEF in the Language Kitchen: A Generative Data Augmentation Leveraging Korean Morpheme Ingredients
Jaehyung Seo, Hyeonseok Moon, Jaewook Lee, Sugyeong Eo, Chanjun Park and Heuiseok Lim 6014

Automatic Debate Evaluation with Argumentation Semantics and Natural Language Argument Graph Networks
Ramon Ruiz-Dolz, Stella Heras and Ana Garcia ................................................................. 6030

Transfer-Free Data-Efficient Multilingual Slot Labeling
Evgenia Razumovskaia, Ivan Vulić and Anna Korhonen ...................................................... 6041

Towards Interpretable Mental Health Analysis with Large Language Models
Kailai Yang, Shaoxiong Ji, Tianlin Zhang, Qianqian Xie, Ziyan Kuang and Sophia Ananiadou 6056

Learning to Rank Generation with Pairwise Partial Rewards
Youngwon Lee, Jinu Lee and Seung-won Hwang ................................................................. 6078

GreedyCAS: Unsupervised Scientific Abstract Segmentation with Normalized Mutual Information
Yingqiang Gao, Jessica Lam, Nianlong Gu and Richard Hahnlooser .................................. 6093

Spoiler Detection as Semantic Text Matching
Ryan Tran, Canwen Xu and Julian McAuley ................................................................. 6109

Multimodal Embodied Plan Prediction Augmented with Synthetic Embodied Dialogue
Aishwarya Padmakumar, Mert Inan, Spandana Gella, Patrick L. Lange and Dilek Hakkani-Tur 6114

GEM: Gestalt Enhanced Markup Language Model for Web Understanding via Render Tree
Zirui Shao, Feiyu Gao, Zhongda Qi, Hangdi Xing, Jiajun Bu, Zhi Yu, Qi Zheng and Xiaozhong Liu ................................................................. 6132

Abstractive Open Information Extraction
Kevin Song Pei, Ishan Jindal and Kevin Chang ................................................................. 6146

CoSyn: Detecting Implicit Hate Speech in Online Conversations Using a Context Synergized Hyperbolic Network
Sreyan Ghosh, Manan Suri, Purva Chiniya, Utkarsh Tyagi, Sonal Kumar and Dinesh Manocha 6159
CLEME: Debiasing Multi-reference Evaluation for Grammatical Error Correction
Jingheng Ye, Yinghui Li, Qingyu Zhou, Yangning Li, Shirong Ma, Hai-Tao Zheng and Ying Shen

Dynamic Top-k Estimation Consolidates Disagreement between Feature Attribution Methods
Jonathan Kamp, Lisa Beinborn and Antske Fokkens

SentiStream: A Co-Training Framework for Adaptive Online Sentiment Analysis in Evolving Data Streams
Yuhao Wu, Karthic Sharma, Chun Wei Seah and Shuhao Zhang

HyperNetwork-based Decoupling to Improve Model Generalization for Few-Shot Relation Extraction
Liang Zhang, Chulun Zhou, Fandong Meng, Jinsong Su, Yidong Chen and Jie Zhou

Solving Hard Analogy Questions with Relation Embedding Chains
Nitesh Kumar and Steven Schockaert

Modeling Empathic Similarity in Personal Narratives
Jocelyn J Shen, Maarten Sap, Pedro Colon-Hernandez, Hae Won Park and Cynthia Breazeal

Tree Prompting: Efficient Task Adaptation without Fine-Tuning
Chandan Singh, John Xavier Morris, Alexander M Rush, Jianfeng Gao and Yuntian Deng

Baize: An Open-Source Chat Model with Parameter-Efficient Tuning on Self-Chat Data
Canwen Xu, Daya Guo, Nan Duan and Julian McAuley

Empathy Intent Drives Empathy Detection
Liting Jiang, Di Wu, Bohui Mao, Yanbing Li and Wushour Slamu

Adaptive End-to-End Metric Learning for Zero-Shot Cross-Domain Slot Filling
Yuanjun Shi, Linzhi Wu and Minglai Shao

BasahaCorpus: An Expanded Linguistic Resource for Readability Assessment in Central Philippine Languages
Joseph Marvin Imperial and Ekaterina Kochmar

ReTAG: Reasoning Aware Table to Analytic Text Generation
Deepanway Ghosal, Preksha Nema and Aravindan Raghuveer

Beyond Factuality: A Comprehensive Evaluation of Large Language Models as Knowledge Generators
Liang Chen, Yang Deng, Yatao Bian, Zeyu Qin, Bingzhe Wu, Tat-Seng Chua and Kam-Fai Wong

Compressing Context to Enhance Inference Efficiency of Large Language Models
Yucheng Li, Bo Dong, Frank Guerin and Chenghua Lin

MoT: Memory-of-Thought Enables ChatGPT to Self-Improve
Xiaonan Li and Xipeng Qiu

4 and 7-bit Labeling for Projective and Non-Projective Dependency Trees
Carlos Gómez-Rodríguez, Diego Roca and David Vilares

Can You Follow Me? Testing Situational Understanding for ChatGPT
Chenghao Yang and Allyson Ettinger

Towards Reliable Misinformation Mitigation: Generalization, Uncertainty, and GPT-4
Kellin Pelrine, Anne Imouza, Camille Thibault, Melina Reksoprodjo, Caleb Alexander Gupta, Joel Christoph, Jean-François Godbout and Reihaneh Rabbany

Ixi
Advancements in Arabic Grammatical Error Detection and Correction: An Empirical Investigation
Bashar Alhafni, Go Inoue, Christian Khairallah and Nizar Habash ................................. 6430

HaluEval: A Large-Scale Hallucination Evaluation Benchmark for Large Language Models
Junyi Li, Xiaoxue Cheng, Xin Zhao, Jian-Yun Nie and Ji-Rong Wen .............................. 6449

Enabling Large Language Models to Generate Text with Citations
Tianyu Gao, Howard Yen, Jiatong Yu and Danqi Chen ............................................. 6465

Revisiting Machine Translation for Cross-lingual Classification
Mikel Artetxe, Vedanuj Goswami, Shruti Bhosale, Angela Fan and Luke Zettlemoyer .... 6489

Pre-Trained Language Models Augmented with Synthetic Scanpaths for Natural Language Understanding
Shuwen Deng, Paul Prasse, David Robert Reich, Tobias Scheffer and Lena Ann Jäger ... 6500

Counting the Bugs in ChatGPT’s Wugs: A Multilingual Investigation into the Morphological Capabilities of a Large Language Model
Leonie Weissweiler, Valentin Hofmann, Anjali Kantharuban, Anna Cai, Ritam Dutt, Amey Hengle, Anubha Kabra, Athvarha Kulkarni, Abhishek Vijayakumar, Haofei Yu, Hinrich Schuetze, Kemal Oflazer and David R Mortensen .................................................. 6508

Adapt in Contexts: Retrieval-Augmented Domain Adaptation via In-Context Learning
Quanyu Long, Wenya Wang and Sinno Jialin Pan ................................................... 6525

Understanding the Inner-workings of Language Models Through Representation Dissimilarity
Davis Brown, Charles Godfrey, Nicholas Konz, Jonathan Tu and Henry Kvinge ........ 6543

Efficient Classification of Long Documents via State-Space Models
Peng Lu, Suyuchen Wang, Mehdi Rezagholizadeh, Bang Liu and Ivan Kobyzev .......... 6559

Dual-Feedback Knowledge Retrieval for Task-Oriented Dialogue Systems
Tianyuan Shi, Liangzhi Li, Zijian Lin, Tao Yang, Xiaojun Quan and Qifan Wang ....... 6566

Construction Artifacts in Metaphor Identification Datasets
Joanne Boisson, Luis Espinosa-Anke and Jose Camacho-Collados ........................... 6581

MAF: Multi-Aspect Feedback for Improving Reasoning in Large Language Models
Deepak Nathani, David Wang, Liangming Pan and William Yang Wang .................. 6591

Granularity Matters: Pathological Graph-driven Cross-modal Alignment for Brain CT Report Generation
Yanzhao Shi, Junzhong Ji, Xiaodan Zhang, Liangqiong Qu and Ying Liu .................. 6617

Enhancing Structured Evidence Extraction for Fact Verification
Zirui Wu, Nan Hu and Yansong Feng ................................................................. 6631

Rethinking Model Selection and Decoding for Keyphrase Generation with Pre-trained Sequence-to-Sequences Models
Di Wu, Wasi Uddin Ahmad and Kai-Wei Chang .................................................... 6642

A Fair and In-Depth Evaluation of Existing End-to-End Entity Linking Systems
Hannah Bast, Matthias Hertel and Natalie Prange ................................................. 6659

A Multi-Task Dataset for Assessing Discourse Coherence in Chinese Essays: Structure, Theme, and Logic Analysis
Hongyi Wu, Xinshu Shen, Man Lan, Shaoguang Mao, Xiaopeng Bai and Yuanbin Wu ... 6673
SKD-NER: Continual Named Entity Recognition via Span-based Knowledge Distillation with Reinforcement Learning
Yi Chen and Liang He .......................................................... 6689

Lifelong Sequence Generation with Dynamic Module Expansion and Adaptation
Chengwei Qin, Chen Chen and Shafiq Joty .................................. 6701

When the Majority is Wrong: Modeling Annotator Disagreement for Subjective Tasks
Eve Fleisig, Rediet Abebe and Dan Klein ..................................... 6715

Lazy-k Decoding: Constrained Decoding for Information Extraction
Arthur Hemmer, Mickael Coustaty, Nicola Bartolo, Jerome Brachat and Jean-marc Ogier . 6727

Personalized Distillation: Empowering Open-Sourced LLMs with Adaptive Learning for Code Generation
Hailin Chen, Amrita Saha, Steven Hoi and Shafiq Joty ....................... 6737

Do Language Models Have a Common Sense regarding Time? Revisiting Temporal Commonsense Reasoning in the Era of Large Language Models
Raghav Jain, Daivik Sojitra, Arkadeep Acharya, Sriparna Saha, Adam Jatowt and Sandipan Dandapat ........................................ 6750

Comparing Styles across Languages
Shreya Havaldar, Matthew Pressimone, Eric Wong and Lyle Ungar ....................... 6775

Event Causality Extraction via Implicit Cause-Effect Interactions
Jintao Liu, Zequn Zhang, Kaiwen Wei, Zhi Guo, Xian Sun, Li Jin and Xiaoyu Li ........ 6792

Evaluation of African American Language Bias in Natural Language Generation
Nicholas Deas, Jessica A Grieser, Shana Kleiner, Desmond U. Patton, Elsbeth Turcan and Kathleen McKeown .......................... 6805

A Systematic Study of Performance Disparities in Multilingual Task-Oriented Dialogue Systems
Songbo Hu, Han Zhou, Moy Yuan, Milan Gritta, Guchun Zhang, Ignacio Iacobacci, Anna Korhonen and Ivan Vulić ........................................ 6825

Cognate Transformer for Automated Phonological Reconstruction and Cognate Reflex Prediction
V.S.D.S.Mahesh Akavarapu and Arnab Bhattacharya ....................... 6852

Inference-Time Policy Adapters (IPA): Tailoring Extreme-Scale LMs without Fine-tuning
Ximing Lu, Faeze Brahman, Peter West, Jaehun Jung, Khyathi Chandu, Abhilasha Ravichander, Prithviraj Anmanabrolu, Liwei Jiang, Sahana Ramnath, Nouha Dziri, Jillian Fisher, Bill Yuchen Lin, Skyler Hallinan, Lianhui Qin, Xiang Ren, Sean Welleck and Yejin Choi .................. 6863

Weakly Supervised Semantic Parsing with Execution-based Spurious Program Filtering
Kang-il Lee, Segwang Kim and Kyomin Jung ................................ 6884

Taxonomy Expansion for Named Entity Recognition
Karthikeyan K, Yogarshi Vyas, Jie Ma, Giovanni Paolini, Neha Anna John, Shuai Wang, Yassine Benajiba, Vittorio Castelli, Dan Roth and Miguel Ballesteros ................ 6895

Rather a Nurse than a Physician - Contrastive Explanations under Investigation
Oliver Eberle, Ilias Chalkidis, Laura Cabello and Stephanie Brandl ........ 6907

EtiCor: Corpus for Analyzing LLMs for Etiquettes
Ashutosh Dwivedi, Pradhyumna Lavania and Ashutosh Modi ........... 6921
An Investigation of LLMs’ Inefficacy in Understanding Converse Relations
Chengwen Qi, Bowen Li, Binyuan Hui, Bailin Wang, Jinyang Li, Jinwang Wu and Yuanjun Laili

Towards Low-Resource Automatic Program Repair with Meta-Learning and Pretrained Language Models
Weishi Wang, Yue Wang, Steven Hoi and Shafiq Joty

ZGUL: Zero-shot Generalization to Unseen Languages using Multi-source Ensembling of Language Adapters
Vipul Kumar Rathore, Rajdeep Dhingra, Parag Singla and Mausam

Log-FGAER: Logic-Guided Fine-Grained Address Entity Recognition from Multi-Turn Spoken Dialogue
Xue Han, Yitong Wang, Qian Hu, Pengwei Hu, Chao Deng and Junlan Feng

Unified Low-Resource Sequence Labeling by Sample-Aware Dynamic Sparse Finetuning
Sarkar Snigdha Sarathi Das, Haoran Ranran Zhang, Peng Shi, Wenpeng Yin and Rui Zhang

On the Representational Capacity of Recurrent Neural Language Models
Franz Nowak, Anej Svete, Li Du and Ryan Cotterell

A Mechanistic Interpretation of Arithmetic Reasoning in Language Models using Causal Mediation Analysis
Alessandro Stolfo, Yonatan Belinkov and Mrinmaya Sachan

Benchmarking and Improving Text-to-SQL Generation under Ambiguity
Adithya Bhaskar, Tushar Tomar, Ashutosh Sathe and Sunita Sarawagi

Non-autoregressive Text Editing with Copy-aware Latent Alignments
Yu Zhang, Yue Zhang, Leyang Cui and Guohong Fu

Translating away Translationese without Parallel Data
Rricha Jalota, Koel Dutta Chowdhury, Cristina España-Bonet and Josef van Genabith

Prompt-Based Monte-Carlo Tree Search for Goal-oriented Dialogue Policy Planning
Xiao Yu, Maximillian Chen and Zhou Yu

UniMath: A Foundational and Multimodal Mathematical Reasoner
Zhenwen Liang, Tianyu Yang, Jipeng Zhang and Xiangliang Zhang

CaseEncoder: A Knowledge-enhanced Pre-trained Model for Legal Case Encoding
Yixiao Ma, Yueyue WU, Weihang Su, Qingyao Ai and Yiqun Liu

HiddenTables and PyQTax: A Cooperative Game and Dataset For TableQA to Ensure Scale and Data Privacy Across a Myriad of Taxonomies
William Watson, Nicole Cho, Tucker Balch and Manuela Veloso

Causal Document-Grounded Dialogue Pre-training
Yingxiu Zhao, Bowen Yu, Bowen Li, Haiyang Yu, Jinyang Li, Chao Wang, Fei Huang, Yongbin Li and Nevin L. Zhang

Accented Speech Recognition With Accent-specific Codebooks
Darshan Deepak Prabhu, Preethi Jyothi, Sriram Ganapathy and Vinit Unni

Linking Surface Facts to Large-Scale Knowledge Graphs
Gorjan Radevski, Kiril Gashteovski, Chia-Chien Hung, Carolin Lawrence and Goran Glavaš
Sentiment Analysis on Streaming User Reviews via Dual-Channel Dynamic Graph Neural Network
Xin Zhang, Linhai Zhang and Deyu Zhou .......................................................... 7208

DUMB: A Dutch Model Benchmark
Wietse de Vries, Martijn Wieling and Malvina Nissim ........................................ 7221

OssCSE: Overcoming Surface Structure Bias in Contrastive Learning for Unsupervised Sentence Embedding
Zhan Shi, Guoyin Wang, Ke Bai, Jiwei Li, Xiang Li, Qingjun Cui, Belinda Zeng, Trishul Chilimbi and Xiaodan Zhu ................................................................. 7242

End-to-End Single-Channel Speaker-Turn Aware Conversational Speech Translation
Juan Pablo Zuluaga-Gomez, Zhaocheng Huang, Xing Niu, Rohit Paturi, Sundararajan Srinivasan, Prashant Mathur, Brian Thompson and Marcello Federico ........................................ 7255

A Fine-Grained Taxonomy of Replies to Hate Speech
Xinchen Yu, Ashley Zhao, Eduardo Blanco and Lingzi Hong .................................... 7275

JointMatch: A Unified Approach for Diverse and Collaborative Pseudo-Labeling to Semi-Supervised Text Classification
Henry Peng Zou and Cornelia Caragea ................................................................... 7290

Simple Temporal Adaptation to Changing Label Sets: Hashtag Prediction via Dense KNN
Niloofar Miresghallah, Nikolai Vogler, Junxian He, Omar Florez, Ahmed El-Kishky and Taylor Berg-Kirkpatrick ................................................................. 7302

Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4
Kent K. Chang, Mackenzie Hanh Cramer, Sandeep Soni and David Bamman ............ 7312

A Study on Accessing Linguistic Information in Pre-Trained Language Models by Using Prompts
Marion Di Marco, Katharina Hämmerl and Alexander Fraser .................................... 7328

CiteBench: A Benchmark for Scientific Citation Text Generation
Martin Funkquist, Ilia Kuznetsov, Yufang Hou and Iryna Gurevych .......................... 7337

From Heuristic to Analytic: Cognitively Motivated Strategies for Coherent Physical Commonsense Reasoning
Zheyuan Zhang, Shane Storks, Fengyuan Hu, Sungryull Sohn, Moontae Lee, Honglak Lee and Joyce Chai ................................................................. 7354

A Challenging Multimodal Video Summary: Simultaneously Extracting and Generating Keyframe-Caption Pairs from Video
Keito Kudo, Haruki Nagasawa, Jun Suzuki and Nobuyuki Shimizu ............................. 7380

Copyright Violations and Large Language Models
Antonia Karamolegkou, Jiaang Li, Li Zhou and Anders Søgaard .............................. 7403

Effects of sub-word segmentation on performance of transformer language models
Jue Hou, Anisia Katinskaia, Anh-Duc Vu and Roman Yangarber ............................. 7413

Symbolic Planning and Code Generation for Grounded Dialogue
Justin T Chiu, Wenting Zhao, Derek Chen, Saujas Vaduguru, Alexander M Rush and Daniel Fried 7426

Universal Self-Adaptive Prompting
Xingchen Wan, Ruoxi Sun, Hootan Nakhhost, Hanjun Dai, Julian Martin Eisenschlos, Sercan O Arik and Tomas Pfister ................................................................. 7437
Somali Information Retrieval Corpus: Bridging the Gap between Query Translation and Dedicated Language Resources
Abdisalam Mahamed Badel, Ting Zhong, Wenxin Tai and Fan Zhou ................................. 7463

Beat LLMs at Their Own Game: Zero-Shot LLM-Generated Text Detection via Querying ChatGPT
Biru Zhu, Lifen Yuan, Ganqu Cui, Yangyi Chen, Chong Fu, Bingxiang He, Yangdong Deng, Zhiyuan Liu, Maosong Sun and Ming Gu .......................................................... 7470

Faithful Model Evaluation for Model-Based Metrics
Qian Hu, Palash Goyal and Rahul Gupta ................................................................. 7484

Content- and Topology-Aware Representation Learning for Scientific Multi-Literature
Kai Zhang, Kaisong Song, Yangyang Kang and Xiaozhong Liu .................................... 7490

Language Model Quality Correlates with Psychometric Predictive Power in Multiple Languages
Ethan Wilcox, Clara Meister, Ryan Cotterell and Tiago Pimentel ............................... 7503

Joint Entity and Relation Extraction with Span Pruning and Hypergraph Neural Networks
Zhaohui Yan, Songlin Yang, Wei Liu and Kewei Tu ............................................... 7512

Have LLMs Advanced Enough? A Challenging Problem Solving Benchmark For Large Language Models
Daman Arora, Himanshu Gaurav Singh and Mausam ............................................... 7527

StrAE: Autoencoding for Pre-Trained Embeddings using Explicit Structure
Mattia Opper, Victor Prokhorov and Siddharth N ................................................... 7544

WiCE: Real-World Entailment for Claims in Wikipedia
Ryo Kamoi, Tanya Goyal, Juan Diego Rodriguez and Greg Durrett ............................ 7561

Natural Disaster Tweets Classification Using Multimodal Data
Mohammad Abdul Basit, Bashir Alam, Zubaida Fatima and Salman Shaikh ............... 7584

On the Challenges of Using Black-Box APIs for Toxicity Evaluation in Research
Luiza Amador Pozzobon, Beyza Ermis, Patrick Lewis and Sara Hooker .................... 7595

RoBoCoP: A Comprehensive ROMance BOrrowing COgnate Package and Benchmark for Multilingual Cognate Identification
Liviu P Dinu, Ana Sabina Uban, Alina Maria Cristea, Anca Daniela Dinu, Ioan-Bogdan Iordache, Simona Georgescu and Laurentiu Zoicas ................................................................. 7610

Instructive Dialogue Summarization with Query Aggregations
Bin Wang, Zhengyuan Liu and Nancy F. Chen ...................................................... 7630

Semantic matching for text classification with complex class descriptions
Brian M De Silva, Kuan-Wen Huang, Gwang Gook Lee, Karen Hovsepian, Yan Xu and Mingwei Shen .................................................................................................................. 7654

MADNet: Maximizing Addressee Deduction Expectation for Multi-Party Conversation Generation
Jia-Chen Gu, Chao-Hong Tan, Caiyuan Chu, Zhen-Hua Ling, Chongyang Tao, Quan Liu and Cong Liu ................................................................. 7681

GLEN: Generative Retrieval via Lexical Index Learning
Sunkyung Lee, Minjin Choi and Jongwuk Lee ......................................................... 7693

Turn-Level Active Learning for Dialogue State Tracking
Zihan Zhang, Meng Fang, Fanghua Ye, Ling Chen and Mohammad-Reza Namazi-Rad .... 7705
ReSee: Responding through Seeing Fine-grained Visual Knowledge in Open-domain Dialogue
Haoqin Tu, Yitong Li, Fei Mi and Zhongliang Yang .................................................... 7720

Modeling Conceptual Attribute Likeness and Domain Inconsistency for Metaphor Detection
Yuan Tian, Nan Xu, Wenji Mao and Daniel Dajun Zeng .................................................. 7736

Referring Image Segmentation via Joint Mask Contextual Embedding Learning and Progressive Alignment Network
Ziling Huang and Shin’ichi Satoh ................................................................. 7753

Shall We Pretrain Autoregressive Language Models with Retrieval? A Comprehensive Study
Boxin Wang, Wei Ping, Peng Xu, Lawrence McAfee, Zihan Liu, Mohammad Shoeybi, Yi Dong, Oleksii Kuchaiev, Bo Li, Chaowei Xiao, Anima Anandkumar and Bryan Catanzaro .................................................. 7763

SCITAB: A Challenging Benchmark for Compositional Reasoning and Claim Verification on Scientific Tables
Xinyuan Lu, Liangming Pan, Qian Liu, Preslav Nakov and Min-Yen Kan ................................ 7787

Training Simultaneous Speech Translation with Robust and Random Wait-k-Tokens Strategy
Linlin Zhang, Kai Fan, Jiajun Bu and Zhongqiang Huang .................................................. 7814

SCENE: Self-Labeled Counterfactuals for Extrapolating to Negative Examples
Deqing Fu, Ameya Godbole and Robin Jia ............................................................. 7832

Enhancing Code-Switching for Cross-lingual SLU: A Unified View of Semantic and Grammatical Coherence
Zhihong Zhu, Xuxin Cheng, Zhiqi Huang, Dongsheng Chen and Yuexian Zou .................. 7849

Task-Agnostic Low-Rank Adapters for Unseen English Dialects
Zedian Xiao, William Barr Held, Yanchen Liu and Diyi Yang ........................................ 7857

Federated Learning of Large Language Models with Parameter-Efficient Prompt Tuning and Adaptive Optimization
Tianshi Che, Ji Liu, Yang Zhou, Jiaxiang Ren, Jiwen Zhou, Victor S. Sheng, Huaiyu Dai and Dejing Dou .......................................................... 7871

TheoremQA: A Theorem-driven Question Answering Dataset
Wenhu Chen, Ming Yin, Max Ku, Pan Lu, Yixin Wan, Xueguang Ma, Jianyu Xu, Xinyi Wang and Tony Xia .............................................................. 7889

Scalable-DSC: A Structural Template Prompt Approach to Scalable Dialogue State Correction
Haoxiang Su, Hongyan Xie, Hao Huang, Shuangyong Song, Ruiyu Fang, Xiaomeng Huang and Sijie Feng .............................................................. 7902

Don’t Trust ChatGPT when your Question is not in English: A Study of Multilingual Abilities and Types of LLMs
Xiang Zhang, Senyu Li, Bradley Hauer, Ning Shi and Grzegorz Kondrak .......................... 7915

M^3Seg: A Maximum-Minimum Mutual Information Paradigm for Unsupervised Topic Segmentation in ASR Transcripts
Ke Wang, Xiutian Zhao, Yanghui Li and Wei Peng .................................................... 7928

Empirical Study of Zero-Shot NER with ChatGPT
Tingyu Xie, Qi Li, Jian Zhang, Yan Zhang, Zuozhu Liu and Hongwei Wang .................. 7935

Automatic Prompt Optimization with “Gradient Descent” and Beam Search
Reid Pryzant, Dan Iter, Jerry Li, Yin Tat Lee, Chenguang Zhu and Michael Zeng ........... 7957

lxvii
Active Retrieval Augmented Generation
Zhengbao Jiang, Frank F. Xu, Luyu Gao, Zhiqing Sun, Qian Liu, Jane Dwivedi-Yu, Yiming Yang, Jamie Callan and Graham Neubig

GD-COMET: A Geo-Diverse Commonsense Inference Model
Mehar Bhatia and Vered Shwartz

Multi-level Adaptive Contrastive Learning for Knowledge Internalization in Dialogue Generation
Chenxu Yang, Zheng Lin, Lanrui Wang, Chong Tian, Liang Pang, Jiangan Li, Qi Hong, Yanan Cao and Weiping Wang

Enhancing Biomedical Lay Summarisation with External Knowledge Graphs
Tomas Goldsack, Zhihao Zhang, Chen Tang, Carolina Scarton and Chenghua Lin

A Diffusion Weighted Graph Framework for New Intent Discovery
Wenkai Shi, Wenbin An, Feng Tian, Qinghua Zheng, Qianying Wang and Ping Chen

A Self-enhancement Multitask Framework for Unsupervised Aspect Category Detection
Thi-Nhung Nguyen, Hoang Ngo, Kiem-Hieu Nguyen and Tuan-Dung Cao

DialCoT Meets PPO: Decomposing and Exploring Reasoning Paths in Smaller Language Models
Chengcheng Han, Xiaowei Du, Che Zhang, Yixin Lian, Xiang Li, Ming Gao and Baoyuan Wang

Recurrent Neural Language Models as Probabilistic Finite-state Automata
Anej Svete and Ryan Cotterell

Revisiting Source Context in Nearest Neighbor Machine Translation
Xuanhong Li, Peng Li and Po Hu

Find-2-Find: Multitask Learning for Anaphora Resolution and Object Localization
Cennet Oguz, Pascal Denis, Emmanuel Vincent, Simon Ostermann and Josef van Genabith

Background Summarization of Event Timelines
Adithya Pratapa, Kevin Small and Markus Dreyer

Superlim: A Swedish Language Understanding Evaluation Benchmark
Aleksandrs Berdicevskis, Gerlof Bouma, Robin Kurtz, Felix Morger, Joey Öhman, Yvonne Adesam, Lars Borin, Dana Dannélius, Markus Forsberg, Tim Isbister, Anna Lindahl, Martin Malmsten, Faton Rekathati, Magnus Sahlgren, Elena Volodina, Love Börjeson, Simon Hengchen and Nina Tahmasebi

Reasoning with Language Model is Planning with World Model
Shibo Hao, Yi Gu, Haodi Ma, Joshua Jiahua Hong, Zhen Wang, Daisy Zhe Wang and Zhiting Hu

LLM-enhanced Self-training for Cross-domain Constituency Parsing
Jianling Li, Meishan Zhang, Peiming Guo, Min Zhang and Yue Zhang

Continual Named Entity Recognition without Catastrophic Forgetting
Duzhen Zhang, Wei Cong, Jiahua Dong, Yahan Yu, Xiuyi Chen, Yonggang Zhang and Zhen Fang

DSI++: Updating Transformer Memory with New Documents
Sanket Vaibhav Mehta, Jai Gupta, Yi Tay, Mostafa Dehghani, Vinh Q. Tran, Jinfeng Rao, Marc Najork, Emma Strubell and Donald Metzler

ixviii
Editing Common Sense in Transformers
Anshita Gupta, Debanjan Mondal, Akshay Krishna Sheshadri, Wenlong Zhao, Xiang Lorraine Li, Sarah Wiegreffe and Niket Tandon ................................................................. 8214

Air-Decoding: Attribute Distribution Reconstruction for Decoding-Time Controllable Text Generation
Tianqi Zhong, Quan Wang, Jingxuan Han, Yongdong Zhang and Zhendong Mao .......... 8233

Homophone Disambiguation Reveals Patterns of Context Mixing in Speech Transformers
Hosein Mohebbi, Grzegorz Chrupała, Willem Zuidema and Afra Alishahi ................. 8249

Retrieval-Generation Alignment for End-to-End Task-Oriented Dialogue System
Weizhou Shen, Yingqi Gao, Canbin Huang, Fanqi Wan, Xiaojun Quan and Wei Bi .... 8261

IfQA: A Dataset for Open-domain Question Answering under Counterfactual Presuppositions
Wenhao Yu, Meng Jiang, Peter Clark and Ashish Sabharwal ................................ 8276

How Do Large Language Models Capture the Ever-changing World Knowledge? A Review of Recent Advances
Zihan Zhang, Meng Fang, Ling Chen, Mohammad-Reza Namazi-Rad and Jun Wang .. 8289

PreWoMe: Exploiting Presuppositions as Working Memory for Long Form Question Answering
Wookje Han, Jinsol Park and Kyungjae Lee .......................................................... 8312

Memorisation Cartography: Mapping out the Memorisation-Generalisation Continuum in Neural Machine Translation
Verna Dankers, Ivan Titov and Dieuwke Hupkes .................................................. 8323

DecipherPref: Analyzing Influential Factors in Human Preference Judgments via GPT-4
Yebowen Hu, Kaiqiang Song, Sangwoo Cho, Xiaoyang Wang, Hassan Foroosh and Fei Liu. 8344

Gender Biases in Automatic Evaluation Metrics for Image Captioning
Haoyi Qiu, Zi-Yi Dou, Tianlu Wang, Asli Celikyilmaz and Nanyun Peng ................. 8358

QA-NatVer: Question Answering for Natural Logic-based Fact Verification
Rami Aly, Marek Strong and Andreas Vlachos ..................................................... 8376

Increasing Probability Mass on Answer Choices Does Not Always Improve Accuracy
Sarah Wiegreffe, Matthew Finlayson, Oyvind Tafjord, Peter Clark and Ashish Sabharwal . 8392

Generating Data for Symbolic Language with Large Language Models
Jiacheng Ye, Chengzu Li, Lingpeng Kong and Tao Yu ......................................... 8418

IDTraffickers: An Authorship Attribution Dataset to link and connect Potential Human-Trafficking Operations on Text Escort Advertisements
Vageesh Kumar Saxena, Benjamin Ashpole, Gijs van Dijck and Gerasimos Spanakis .... 8444

Evaluating Bias and Fairness in Gender-Neutral Pretrained Vision-and-Language Models
Laura Cabello, Emanuele Bugliarello, Stephanie Brandl and Desmond Elliott ............ 8465

Improving Dialogue Discourse Parsing via Reply-to Structures of Addressee Recognition
Yaxin Fan, Feng Jiang, Peifeng Li, Fang Kong and Qiaoming Zhu ......................... 8484

Improving Language Models’ Meaning Understanding and Consistency by Learning Conceptual Roles from Dictionary
Myeongjun Erik Jang and Thomas Lukasiewicz .................................................. 8496
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event-Location Tracking in Narratives: A Case Study on Holocaust Testimonies</td>
<td>Eitan Wagner, Renana Keydar and Omri Abend</td>
<td>8789</td>
</tr>
<tr>
<td>Dialogizer: Context-aware Conversational-QA Dataset Generation from Textual Sources</td>
<td>Yerin Hwang, Yongil Kim, Hyunkyung Bae, Hwanhee Lee, Jeesoo Bang and Kyomin Jung</td>
<td>8806</td>
</tr>
<tr>
<td>Learning to Predict Task Transferability via Soft Prompt</td>
<td>Lingyun Feng</td>
<td>8829</td>
</tr>
<tr>
<td>Chain-of-Questions Training with Latent Answers for Robust Multistep Question Answering</td>
<td>Wang Zhu, Jesse Thomason and Robin Jia</td>
<td>8845</td>
</tr>
<tr>
<td>Mirror: A Universal Framework for Various Information Extraction Tasks</td>
<td>Tong Zhu, Junfei Ren, Zijian Yu, Mengsong Wu, Guoliang Zhang, Xiaoye Qu, Wenliang Chen, Zhefeng Wang, Baoxing Huai and Min Zhang</td>
<td>8861</td>
</tr>
<tr>
<td>“Mistakes Help Us Grow”: Facilitating and Evaluating Growth Mindset Supportive Language in Classrooms</td>
<td>Kunal Handa, Margarett Clapper, Jessica Boyle, Rose E Wang, Diyi Yang, David Yeager and Dorottya Demszky</td>
<td>8877</td>
</tr>
<tr>
<td>Unnatural Error Correction: GPT-4 Can Almost Perfectly Handle Unnatural Scrambled Text</td>
<td>Qi Cao, Takeshi Kojima, Yutaka Matsuo and Yusuke Iwasawa</td>
<td>8898</td>
</tr>
<tr>
<td>Detecting and Mitigating Hallucinations in Multilingual Summarisation</td>
<td>Yifu Qiu, Yifah Ziser, Anna Korhonen, Edoardo Ponti and Shay B. Cohen</td>
<td>8914</td>
</tr>
<tr>
<td>Exploring Linguistic Probes for Morphological Inflection</td>
<td>Jordan Kodner, Salam Khalifa and Sarah Ruth Brogden Payne</td>
<td>8933</td>
</tr>
<tr>
<td>AMR Parsing with Causal Hierarchical Attention and Pointers</td>
<td>Chao Lou and Kewei Tu</td>
<td>8942</td>
</tr>
<tr>
<td>FLatS: Principled Out-of-Distribution Detection with Feature-Based Likelihood Ratio Score</td>
<td>Haowei Lin and Yuntian Gu</td>
<td>8956</td>
</tr>
<tr>
<td>Self-Evolution Learning for Mixup: Enhance Data Augmentation on Few-Shot Text Classification Tasks</td>
<td>Haoqi Zheng, Qihuang Zhong, Liang Ding, Zhiliang Tian, Xin Niu, Changjian Wang, Dongsheng Li and Dacheng Tao</td>
<td>8964</td>
</tr>
<tr>
<td>IC3: Image Captioning by Committee Consensus</td>
<td>David Chan, Austin Myers, Sudheendra Vijayanarasimhan, David A Ross and John Canny</td>
<td>8975</td>
</tr>
<tr>
<td>SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models</td>
<td>Potsawee Manakul, Adian Liusie and Mark Gales</td>
<td>9004</td>
</tr>
<tr>
<td>Fair Without Leveling Down: A New Intersectional Fairness Definition</td>
<td>Gaurav Maheshwari, Aurélien Bellet, Pascal Denis and Mikaela Keller</td>
<td>9018</td>
</tr>
<tr>
<td>Revisiting Instruction Fine-tuned Model Evaluation to Guide Industrial Applications</td>
<td>Manuel Faysse, Gautier Viaud, Céline Hudelot and Pierre Colombo</td>
<td>9033</td>
</tr>
<tr>
<td>CLAD-ST: Contrastive Learning with Adversarial Data for Robust Speech Translation</td>
<td>Sathish Reddy Indurthi, Shamil Chollampatt, Ravi Agrawal and Marco Turchi</td>
<td>9049</td>
</tr>
</tbody>
</table>
Xurui Li, Yue Qin, Rui Zhu, Tianqitian Lin, Yongming Fan, Yangyang Kang, Kaisong Song, Fubang Zhao, Changlong Sun, Haixu Tang and Xiaozhong Liu ................................. 9304

Centering the Margins: Outlier-Based Identification of Harmed Populations in Toxicity Detection
Vyoma Raman, Eve Fleisig and Dan Klein .................................................. 9316

Describe Me an Auklet: Generating Grounded Perceptual Category Descriptions
Bill Noble and Nikolai Ilinykh ................................................................. 9330

Revisiting Automated Topic Model Evaluation with Large Language Models
Dominik Stammbach, Vilém Zouhar, Alexander Hoyle, Mrinmaya Sachan and Elliott Ash .......................... 9348

ORCHID: A Chinese Debate Corpus for Target-Independent Stance Detection and Argumentative Dialogue Summarization
Xiutian Zhao, Ke Wang and Wei Peng ...................................................... 9358

On the Benefits of Learning to Route in Mixture-of-Experts Models
Nishanth Dikkala, Nikhil Ghosh, Raghu Meka, Rina Panigrahy, Nikhil Vyas and Xin Wang 9376

SEAHORSE: A Multilingual, Multifaceted Dataset for Summarization Evaluation
Elizabeth Clark, Shruti Rijhwani, Sebastian Gehrmann, Joshua Maynez, Roeie Aharoni, Vitaly Nikolaev, Thibault Sellam, Aditya Siddhant, Dipanjan Das and Ankur P Parikh ...................... 9397

Query2doc: Query Expansion with Large Language Models
Liang Wang, Nan Yang and Furu Wei .......................................................... 9414

We Need to Talk About Reproducibility in NLP Model Comparison
Yan Xue, Xuefei Cao, Xingli Yang, Yu Wang, Ruibo Wang and Jihong Li .................. 9424

Explore-Instruct: Enhancing Domain-Specific Instruction Coverage through Active Exploration
Fanqi Wan, Xinting Huang, Tao Yang, Xiaojun Quan, Wei Bi and Shuming Shi .................. 9435

Practical Computational Power of Linear Transformers and Their Recurrent and Self-Referential Extensions
Kazuki Irie, Róbert Csordás and Jürgen Schmidhuber ................................ 9455

InterFair: Debiasing with Natural Language Feedback for Fair Interpretable Predictions
Bodhisattwa Prasad Majumder, Zexue He and Julian McAuley ..................... 9466

Just Adjust One Prompt: Enhancing In-Context Dialogue Scoring via Constructing the Optimal Subgraph of Demonstrations and Prompts
Jiashu Pu, Ling Cheng, Lu Fan, Tangjie Lv and Rongsheng Zhang .................. 9472

Multilingual estimation of political-party positioning: From label aggregation to long-input Transformers
Dmitry Nikolaev, Tanise Ceron and Sebastian Padó ................................... 9497

ART: rule bAsed futuRe-inference deducTion
Mengze Li, Tianqi Zhao, Bai Jionghao, Baoyi He, Jiaxu Miao, Wei Ji, Zheqi Lv, Zhou Zhao, Shengyu Zhang, Wenqiao Zhang and Fei Wu ................................. 9512

EpiK-Eval: Evaluation for Language Models as Epistemic Models
Gabriele Prato, Jerry Huang, Prasanna Parthasarathi, Shagun Sodhani and Sarath Chandar 9523

lxxiii
From Dissonance to Insights: Dissecting Disagreements in Rationale Construction for Case Outcome Classification
Shanshan Xu, Santosh T.Y.S.S, Oana Ichim, Isabella Risini, Barbara Plank and Matthias Grabmair
9558

On Bilingual Lexicon Induction with Large Language Models
Yaoyiran Li, Anna Korhonen and Ivan Vulić ................................................................. 9577

Statistical Depth for Ranking and Characterizing Transformer-Based Text Embeddings
Parker Seegmiller and Sarah Masud Preum ................................................................. 9600

CRaSh: Clustering, Removing, and Sharing Enhance Fine-tuning without Full Large Language Model
Kaiyan Zhang, Ning Ding, Biqing Qi, Xuekai Zhu, Xinwei Long and Bowen Zhou ...... 9612

From Multilingual Complexity to Emotional Clarity: Leveraging Commonsense to Unveil Emotions in Code-Mixed Dialogues
Shivani Kumar, Ramaneswaran S, Md Shad Akhtar and Tanmoy Chakraborty ....... 9638

Large Language Models are biased to overestimate profundness
Eugenio Herrera-Berg, Tomás Vergara Browne, Pablo León-Villagrá, Marc-Lluís Vives and Christian Buc Calderon .......................................................... 9653

SummEdits: Measuring LLM Ability at Factual Reasoning Through The Lens of Summarization
Philippe Laban, Wojciech Maciej Kryscinski, Divyansh Agarwal, Alexander Fabbri, Caiming Xiong, Shafiq Joty and Chien-Sheng Wu .......................................................... 9662

DIVE: Towards Descriptive and Diverse Visual Commonsense Generation
Jun-Hyung Park, Hyuntae Park, Youjin Kang, Eojin Jeon and SangKeun Lee ........... 9677

Towards Conceptualization of “Fair Explanation”: Disparate Impacts of anti-Asian Hate Speech Explanations on Content Moderators
Tin Trung Nguyen, Jiannan Xu, Aayushi Roy, Hal Daumé III and Marine Carpuat ...... 9696

Bridging Background Knowledge Gaps in Translation with Automatic Explicitation
HyoJung Han, Jordan Lee Boyd-Graber and Marine Carpuat .................................. 9718

A Quality-based Syntactic Template Retriever for Syntactically-Controlled Paraphrase Generation
Xue Zhang, Songming Zhang, Yunlong Liang, Yufeng Chen, Jian Liu, Wenjuan Han and Jinan Xu .................................................................................................................. 9736

Beyond Shared Vocabulary: Increasing Representational Word Similarities across Languages for Multilingual Machine Translation
Di Wu and Christof Monz ......................................................................................... 9749

Quantifying the redundancy between prosody and text
Lukas Wolf, Tiago Pimentel, Evelina Fedorenko, Ryan Cotterell, Alex Warstadt, Ethan Wilcox and Tamar I Regev .............................................................. 9765

CRoW: Benchmarking Commonsense Reasoning in Real-World Tasks
Mete Ismayilzada, Debjit Paul, Syrielle Montariol, Mor Geva and Antoine Bosselut .... 9785

A Video Is Worth 4096 Tokens: Verbalize Story Videos To Understand Them In Zero Shot
Aanisha Bhattacharyya, Yaman K Singla, Balaji Krishnamurthy, Rajiv Ratn Shah and Changyou Chen .................................................................................................................. 9822

Label Words are Anchors: An Information Flow Perspective for Understanding In-Context Learning
Lean Wang, Lei Li, Damai Dai, Deli Chen, Hao Zhou, Fandong Meng, Jie Zhou and Xu Sun
PromptST: Abstract Prompt Learning for End-to-End Speech Translation
Tengfei Yu, Liang Ding, Xuebo Liu, Kehai Chen, Meishan Zhang, Dacheng Tao and Min Zhang

Text Rendering Strategies for Pixel Language Models
Jonas F. Lotz, Elizabeth Salesky, Phillip Rust and Desmond Elliott

APoLo: Unified Adapter and Prompt Learning for Vision Language Models
Sanjoy Chowdhury, Sayan Nag and Dinesh Manocha

SAMRank: Unsupervised Keyphrase Extraction using Self-Attention Map in BERT and GPT-2
Byungha Kang and Youhyun Shin

Contrastive Learning for Inference in Dialogue
Etsuko Ishii, Yan Xu, Bryan Wilie, Ziwei Ji, Holy Lovenia, Willy Chung and Pascale Fung

Editing Large Language Models: Problems, Methods, and Opportunities
Yunzhi Yao, Peng Wang, Bozhong Tian, Siyuan Cheng, Zhiou Li, Shumin Deng, Huajun Chen and Ningyu Zhang

MarkQA: A large scale KBQA dataset with numerical reasoning
Xiang Huang, Sitao Cheng, Yuheng Bao, Shanshan Huang and Yuzhong Qu

Comparing Biases and the Impact of Multilingual Training across Multiple Languages
Sharon Levy, Neha Anna John, Ling Liu, Yogarshi Vyas, Jie Ma, Yoshinari Fujinuma, Miguel Ballesteros, Vittorio Castelli and Dan Roth

HutCRS: Hierarchical User-Interest Tracking for Conversational Recommender System
Mingjie Qian, Yongsen Zheng, Jinghui Qin and Liang Lin

Large Language Models Meet Open-World Intent Discovery and Recognition: An Evaluation of ChatGPT
Xiaoshuai Song, Keqing He, Pei Wang, Guanting Dong, Yutao Mou, Jingang Wang, Yunseun Xian, Xunliang Cai and Weiran Xu

The Distributional Hypothesis Does Not Fully Explain the Benefits of Masked Language Model Pretraining
Ting-Rui Chiang and Dani Yogatama

Simple and Effective Input Reformulations for Translation
Brian Yu, Hansen Lillemark and Kurt Keutzer

Pointwise Mutual Information Based Metric and Decoding Strategy for Faithful Generation in Document Grounded Dialogs
Yatin Nandwani, Vineet Kumar, Dinesh Raghu, Sachindra Joshi and Luis A. Lastras

The ACL OCL Corpus: Advancing Open Science in Computational Linguistics
Shaurya Rohatgi, Yanxia Qin, Benjamin Aw, Niranjana Anand Unnithan and Min-Yen Kan

Using Artificial French Data to Understand the Emergence of Gender Bias in Transformer Language Models
Lina Conti and Guillaume Wisniewski

Learning to Rank Context for Named Entity Recognition Using a Synthetic Dataset
Arthur Amalvy, Vincent Labatut and Richard Dufour
Improving Diversity of Demographic Representation in Large Language Models via Collective-Critiques and Self-Voting
Preethi Lahoti, Nicholas Blumm, Xiao Ma, Raghavendra Kotikalapudi, Sahitya Potluri, Qijun Tan, Hansa Srinivasan, Ben Packer, Ahmad Beirami, Alex Beutel and Jilin Chen

Hidding the Ghostwriters: An Adversarial Evaluation of AI-Generated Student Essay Detection
Xinlin Peng, Ying Zhou, Ben He, Le Sun and Yingfei Sun

Contextual Interaction for Argument Post Quality Assessment
Yiran Wang, Xuanang Chen, Ben He and Le Sun

Pre-training Intent-Aware Encoders for Zero- and Few-Shot Intent Classification
Mujeen Sung, James Gung, Elman Mansimov, Nikolaos Pappas, Raphael Shu, Salvatore Romeo, Yi Zhang and Vittorio Castelli

Synthetic Data Generation with Large Language Models for Text Classification: Potential and Limitations
Zhuoyan Li, Hangxiao Zhu, Zhuoran Lu and Ming Yin

GazeVQA: A Video Question Answering Dataset for Multiview Eye-Gaze Task-Oriented Collaborations
Muhammet Furkan Ilaslan, Chenan Song, Joya Chen, Difei Gao, Weixian Lei, Qianli Xu, Joo Hwee Lim and Mike Zheng Shou

People Make Better Edits: Measuring the Efficacy of LLM-Generated Counterfactually Augmented Data for Harmful Language Detection
Indira Sen, Dennis Assenmacher, Mattia Samory, Isabelle Augenstein, Wil Aalst and Claudia Wagner

Unraveling Feature Extraction Mechanisms in Neural Networks
Xiaobing Sun, Jiaxi Li and Wei Lu

CAPSTONE: Curriculum Sampling for Dense Retrieval with Document Expansion
Xingwei He, Yeyun Gong, A-Long Jin, Hang Zhang, Anlei Dong, Jian Jiao, Siu Ming Yiu and Nan Duan

Balance Act: Mitigating Hubness in Cross-Modal Retrieval with Query and Gallery Banks
Yimu Wang, Xiangru Jian and Bo Xue

E-CORE: Emotion Correlation Enhanced Empathetic Dialogue Generation
Fengyi Fu, Lei Zhang, Quan Wang and Zhendong Mao

What do Deck Chairs and Sun Hats Have in Common? Uncovering Shared Properties in Large Concept Vocabularies
Amit Gajbhiye, Zied Bouraoui, Na Li, Usashi Chatterjee, Luis Espinosa-Anke and Steven Schockaert

ALDi: Quantifying the Arabic Level of Dialectness of Text
Amr Keleg, Sharon Goldwater and Walid Magdy

3DRP-Net: 3D Relative Position-aware Network for 3D Visual Grounding
Zehan Wang, Haifeng Huang, Yang Zhao, Linjun Li, Xize Cheng, Yichen Zhu, Aoxiong Yin and Zhou Zhao

Goal-Driven Explainable Clustering via Language Descriptions
Zihan Wang, Jingbo Shang and Ruiqi Zhong
Cross-Lingual Consistency of Factual Knowledge in Multilingual Language Models
Jirui Qi, Raquel Fernández and Arianna Bisazza .................................................. 10650

Learning from Mistakes via Cooperative Study Assistant for Large Language Models
Danqing Wang and Lei Li ................................................................. 10667

Bridging the Digital Divide: Performance Variation across Socio-Economic Factors in Vision-Language Models
Joan Nwatu, Oana Ignat and Rada Mihalcea .................................................. 10686

Conceptor-Aided Debiasing of Large Language Models
Li S. Yifei, Lyle Ungar and João Sedoc .......................................................... 10703

AMR Parsing is Far from Solved: GrAPES, the Granular AMR Parsing Evaluation Suite
Jonas Groschwitz, Shay B Cohen, Lucia Donatelli and Meaghan Fowlie ............. 10728

Rethinking and Improving Multi-task Learning for End-to-end Speech Translation
Yuhao Zhang, Chen Xu, Bei Li, Hao Chen, Tong Xiao, Chunliang Zhang and Jingbo Zhu. 10753

AD-NLP: A Benchmark for Anomaly Detection in Natural Language Processing
Matei Bejan, Andrei Manolache and Marius Popescu ........................................ 10766

Enhancing the Ranking Context of Dense Retrieval through Reciprocal Nearest Neighbors
George Zerveas, Navid Rekabsaz and Carsten Eickhoff .................................... 10779

Cross-Lingual Cross-Target Stance Detection with Dual Knowledge Distillation Framework
Ruike Zhang, Hanxuan Yang and Wenji Mao .................................................. 10804

PRESTO: A Multilingual Dataset for Parsing Realistic Task-Oriented Dialogs
Rahul Goel, Waleed Ammar, Aditya Gupta, Siddharth Vashishta, Motoki Sano, Faiz Surani, Max Chang, Hyunjong Choe, David Greene, Chuan He, Rattima Nitisaroj, Anna Trukhina, Shachi Paul, Pararth Shah, Rushin Shah and Zhou Yu .................................................. 10820

An Iteratively Parallel Generation Method with the Pre-Filling Strategy for Document-level Event Extraction
Guanhua Huang, Runxin Xu, Ying Zeng, Jiaze Chen, Zhouwang Yang and Weinan E . 10834

CoMPosT: Characterizing and Evaluating Caricature in LLM Simulations
Myra Cheng, Tiziano Piccardi and Diyi Yang .................................................. 10853

Chen Huang, Peixin Qin, Wenqiang Lei and Jiancheng Lv .................................. 10876

BERTie Bott’s Every Flavor Labels: A Tasty Introduction to Semantic Role Labeling for Galician
Micaella Bruton and Meriem Beloucif .......................................................... 10892

Program Translation via Code Distillation
Yufan Huang, Mengnan Qi, Yongqiang Yao, Maoquan Wang, Bin Gu, Colin Clement and Neel Sundaresan ................................................................. 10903

FaMeSumm: Investigating and Improving Faithfulness of Medical Summarization
Nan Zhang, Yusen Zhang, Wu Guo, Prasenjit Mitra and Rui Zhang ................. 10915

Grammar-Constrained Decoding for Structured NLP Tasks without Finetuning
Saibo Geng, Martin Josifoski, Maxime Peyrard and Robert West ...................... 10932
Systematic word meta-sense extension
Lei Yu .................................................................................................................. 10953

Ziang Xiao, Susu Zhang, Vivian Lai and Q. Vera Liao .................................................. 10967

Revisiting the Knowledge Injection Frameworks
Peng Fu, Yiming Zhang, Haobo Wang, Weikang Qiu and Junbo Zhao ......................... 10983

We Are What We Repeatedly Do: Inducing and Deploying Habitual Schemas in Persona-Based Responses
Benjamin Kane and Lenhart K. Schubert ................................................................. 10998

Zero-shot Faithfulness Evaluation for Text Summarization with Foundation Language Model
Qi Jia, Siyu Ren, Yizhu Liu and Kenny Q. Zhu ......................................................... 11017

TaskWeb: Selecting Better Source Tasks for Multi-task NLP
Joongwon Kim, Akari Asai, Gabriel Ilharco and Hannaneh Hajishirzi ......................... 11032

Improving Bias Mitigation through Bias Experts in Natural Language Understanding
Eojin Jeon, Mingyu Lee, Juhyeong Park, Yeachan Kim, Wing-Lam Mok and SangKeun Lee 11053

Semi-supervised multimodal coreference resolution in image narrations
Arushi Goel, Basura Fernando, Frank Keller and Hakan Bilen ........................................... 11067

A Predictive Factor Analysis of Social Biases and Task-Performance in Pretrained Masked Language Models
Yi Zhou, Jose Camacho-Collados and Danushka Bollegala .................................................... 11082

Argument-based Detection and Classification of Fallacies in Political Debates
Pierpaolo Goffredo, Mariana Chaves Espinoza, Serena Villata and Elena Cabrio .................. 11101

Collaborative Generative AI: Integrating GPT-k for Efficient Editing in Text-to-Image Generation
Wanrong Zhu, Xinyi Wang, Yujie Lu, Tsu-Jui Fu, Xin Eric Wang, Miguel Eckstein and William Yang Wang .................................................................................................................. 11113

SpEL: Structured Prediction for Entity Linking
Hassan Shavarani and Anoop Sarkar .................................................................................. 11123

Architectural Sweet Spots for Modeling Human Label Variation by the Example of Argument Quality: It’s Best to Relate Perspectives!
Philipp Heinisch, Matthias Orlikowski, Julia Romberg and Philipp Cimiano .................. 11138

Explicit Planning Helps Language Models in Logical Reasoning
Hongyu Zhao, Kangrui Wang, Mo Yu and Hongyuan Mei ................................................ 11155

clembench: Using Game Play to Evaluate Chat-Optimized Language Models as Conversational Agents
Kranti Chalamalasetti, Jana Götte, Sherzod Hakimov, Brielen Madureira, Philipp Sadler and David Schlangen .......................................................................................................................... 11174

Explaining with Contrastive Phrasal Highlighting: A Case Study in Assisting Humans to Detect Translation Differences
Eleftheria Briakou, Navita Goyal and Marine Carpuat .................................................................. 11220

Polyglot or Not? Measuring Multilingual Encyclopedic Knowledge in Foundation Models
Tim Schott, Daniel Ryan Furman and Shreshta Bhat .......................................................... 11238

Ixxix
Anchoring Fine-tuning of Sentence Transformer with Semantic Label Information for Efficient Truly Few-shot Classification
Amalie Brogaard Pauli, Leon Derczynski and Ira Assent ................................. 11254

UDAPDR: Unsupervised Domain Adaptation via LLM Prompting and Distillation of Rerankers
Jon Saad-Falcon, Omar Khattab, Keshav Santhanam, Radu Florian, Martin Franz, Salim Roukos, Avirup Sil, Md Arafat Sultan and Christopher Potts ........................................... 11265

TATA: Stance Detection via Topic-Agnostic and Topic-Aware Embeddings
Hans William Alexander Hanley and Zakir Durumeric ................................. 11280

Data Similarity is Not Enough to Explain Language Model Performance
Gregory Yauney, Emily Reif and David Mimno ................................. 11295

Zero-shot Sharpness-Aware Quantization for Pre-trained Language Models
Miaoxi Zhu, Qihuang Zhong, Li Shen, Liang Ding, Juhua Liu, Bo Du and Dacheng Tao ................................. 11305

Deciphering Stereotypes in Pre-Trained Language Models
Weicheng Ma, Henry Scheible, Brian C Wang, Goutham Veeramachaneni, Pratim Chowdhary, Alan Sun, Andrew Koulougeorge, Lili Wang, Diyi Yang and Soroush Vosoughi ................................. 11328

An Integrative Survey on Mental Health Conversational Agents to Bridge Computer Science and Medical Perspectives
Young Min Cho, Sunny Rai, Lyle Ungar, João Sedoc and Sharath Chandra Guntuku ................................. 11346

Do LLMs Understand Social Knowledge? Evaluating the Sociability of Large Language Models with SocKET Benchmark
Minje Choi, Jiaxin Pei, Sagar Kumar, Chang Shu and David Jurgens ................................. 11370

Interventional Rationalization
Linan Yue, Qi Liu, Li Wang, Yanqing An, Yichao Du and Zhenya Huang ................................. 11404

Don't Take This Out of Context!: On the Need for Contextual Models and Evaluations for Stylistic Rewriting
Akhila Yerukola, Xuhui Zhou, Elizabeth Clark and Maarten Sap ................................. 11419

Axiomatic Preference Modeling for Longform Question Answering
Corby Rosset, Guoqing Zheng, Victor Dibia, Ahmed Hassan Awadallah and Paul N. Bennett 11445

Countering Misinformation via Emotional Response Generation
Daniel Russo, Shane Peter Kaszefski-Yaschuk, Jacopo Staiano and Marco Guerini ................................. 11476

Seq2seq is All You Need for Coreference Resolution
Wenzheng Zhang, Sam Wiseman and Karl Stratos ................................. 11493

Integrating Language Models into Direct Speech Translation: An Inference-Time Solution to Control Gender Inflection
Dennis Fucci, Marco Gaido, Sara Papi, Mauro Cettolo, Matteo Negri and Luisa Bentivogli 11505

StoryAnalogy: Deriving Story-level Analogies from Large Language Models to Unlock Analogical Understanding
Cheng Jiayang, Lin Qiu, Tsz Ho Chan, Tianqing Fang, Weiqi Wang, Chunkit Chan, Dongyu Ru, Qipeng Guo, Hongming Zhang, Yangqiu Song, Yue Zhang and Zheng Zhang ................................. 11518

Beyond Detection: A Defend-and-Summarize Strategy for Robust and Interpretable Rumor Analysis on Social Media
Yi-Ting Chang, Yun-Zhu Song, Yi-Syuan Chen and Hong-Han Shuai ................................. 11538

lxxx
Crystal: Introspective Reasoners Reinforced with Self-Feedback
Jiacheng Liu, Ramakanth Pasunuru, Hannaneh Hajishirzi, Yejin Choi and Asli Celikyilmaz

DiffS2UT: A Semantic Preserving Diffusion Model for Textless Direct Speech-to-Speech Translation
Yongxin Zhu, Zhujin Gao, Xinyuan Zhou, Ye Zhongyi and Linli Xu

BioFEG: Generate Latent Features for Biomedical Entity Linking
Xuhui Sui, Ying Zhang, Xiangrui Cai, Kehui Song, Baohang Zhou, Xiaojie Yuan and Wensheng Zhang

TRIGO: Benchmarking Formal Mathematical Proof Reduction for Generative Language Models
Jing Xiong, Jianhao Shen, Ye Yuan, Haiming Wang, Yichun Yin, Zhengying Liu, Lin Li, Zhijiang Guo, Qingxing Cao, Yinya Huang, Chuanyang Zheng, Xiaodan Liang, Ming Zhang and Qun Liu

Physician Detection of Clinical Harm in Machine Translation: Quality Estimation Aids in Reliance and Backtranslation Identifies Critical Errors
Nikita Mehandru, Sweta Agrawal, Yimin Xiao, Ge Gao, Elaine C Khoong, Marine Carpuat and Niloufar Salehi

Vicarious Offense and Noise Audit of Offensive Speech Classifiers: Unifying Human and Machine Disagreement on What is Offensive
Tharindu Cyril Weerasooriya, Sujan Dutta, Tharindu Ranasinghe, Marcos Zampieri, Christopher M Homan and Ashiqur R. KhudaBukhsh

Generating Summaries with Controllable Readability Levels
Leonardo F. R. Ribeiro, Mohit Bansal and Markus Dreyer

mAggretriever: A Simple yet Effective Approach to Zero-Shot Multilingual Dense Retrieval
Sheng-Chieh Lin, Amin Ahmad and Jimmy Lin

CodeFusion: A Pre-trained Diffusion Model for Code Generation
Mukul Singh, José Cambronero, Sumit Gulwani, Vu Le, Carina Suzana Negreanu and Gust Verbruggen

CESAR: Automatic Induction of Compositional Instructions for Multi-turn Dialogs
Taha Aksu, Devamanyu Hazarika, Shikib Mehri, Seokhwan Kim, Dilek Hakkani-Tur, Yang Liu and Mahdi Namazifar

VECHR: A Dataset for Explainable and Robust Classification of Vulnerability Type in the European Court of Human Rights
Shanshan Xu, Leon Stauffer, Santosh T.Y.S.S, Oana Ichim, Corina Heri and Matthias Grabmair

ACQUIRED: A Dataset for Answering Counterfactual Questions In Real-Life Videos
Te-Lin Wu, Zi-Yi Dou, Qingyuan Hu, Yu Hou, Nischal Reddy Chandra, Marjorie Freedman, Ralph M. Weischedel and Nanyun Peng

From Parse-Execute to Parse-Execute-Refine: Improving Semantic Parser for Complex Question Answering over Knowledge Base
Wangzhen Guo, Linyin Luo, Hanjiang Lai and Jian Yin

Reward-Augmented Decoding: Efficient Controlled Text Generation With a Unidirectional Reward Model
Haikang Deng and Colin Raffel
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CORE: A Few-Shot Company Relation Classification Dataset for Robust Domain Adaptation.</td>
<td>Philipp Borchert, Jochen De Weerdt, Kristof Coussenent, Arno De Caigny and Marie-Francine Moens</td>
<td>11792</td>
</tr>
<tr>
<td>Models See Hallucinations: Evaluating the Factuality in Video Captioning</td>
<td>Hui Liu and Xiaojun Wan</td>
<td>11807</td>
</tr>
<tr>
<td>Back Transcription as a Method for Evaluating Robustness of Natural Language Understanding Models to Speech Recognition Errors</td>
<td>Marek Kubis, Paweł Marek Skórzewski, Marcin Sowański and Tomasz Ziętkiewicz</td>
<td>11824</td>
</tr>
<tr>
<td>Cabbage Sweeter than Cake? Analysing the Potential of Large Language Models for Learning Conceptual Spaces</td>
<td>Usashi Chatterjee, Amit Gajbhiye and Steven Schockaert</td>
<td>11836</td>
</tr>
<tr>
<td>Can Language Models Understand Physical Concepts?</td>
<td>Lei Li, Jingjing Xu, Qingxiu Dong, Ce Zheng, Xu Sun, Lingpeng Kong and Qi Liu</td>
<td>11843</td>
</tr>
<tr>
<td>SPT: Learning to Selectively Insert Prompts for Better Prompt Tuning</td>
<td>Wei Zhu and Ming Tan</td>
<td>11862</td>
</tr>
<tr>
<td>Once Upon a Time in Graph: Relative-Time Pretraining for Complex Temporal Reasoning</td>
<td>Sen Yang, Xin Li, Lidong Bing and Wai Lam</td>
<td>11879</td>
</tr>
<tr>
<td>Expository Text Generation: Imitate, Retrieve, Paraphrase</td>
<td>Nishant Balepur, Jie Huang and Kevin Chang</td>
<td>11896</td>
</tr>
<tr>
<td>Large-scale similarity search with Optimal Transport</td>
<td>Cléa Mehnia Laouar, Yuki Takezawa and Makoto Yamada</td>
<td>11920</td>
</tr>
<tr>
<td>Enhancing Textbooks with Visuals from the Web for Improved Learning</td>
<td>Janvijay Singh, Vilém Zouhar and Mrinmaya Sachan</td>
<td>11931</td>
</tr>
<tr>
<td>Continual Event Extraction with Semantic Confusion Rectification</td>
<td>Zitao Wang, Xinyi Wang and Wei Hu</td>
<td>11945</td>
</tr>
<tr>
<td>An Empirical Study of Translation Hypothesis Ensembling with Large Language Models</td>
<td>António Farinhas, José G. C. de Souza and Andre Martins</td>
<td>11956</td>
</tr>
<tr>
<td>FedTherapist: Mental Health Monitoring with User-Generated Linguistic Expressions on Smartphones via Federated Learning</td>
<td>Jaemin Shin, Hyungjun Yoon, Seunghoo Lee, Sungjoo Park, Yunxin Liu, Jinho D. Choi and Sung-Ju Lee</td>
<td>11971</td>
</tr>
<tr>
<td>Visually-Situated Natural Language Understanding with Contrastive Reading Model and Frozen Large Language Models</td>
<td>Geewook Kim, Hodong Lee, Daeehee Kim, Haeji Jung, Sanghee Park, Yoonsik Kim, Sangdoo Yun, Taeho Kil, Bado Lee and Seunghyun Park</td>
<td>11989</td>
</tr>
<tr>
<td>Continual Learning for Multilingual Neural Machine Translation via Dual Importance-based Model Division</td>
<td>Junpeng Liu, Kaiyu Huang, Hao Yu, Jiuyi Li, Jinsong Su and Degen Huang</td>
<td>12011</td>
</tr>
<tr>
<td>SimCSE++: Improving Contrastive Learning for Sentence Embeddings from Two Perspectives</td>
<td>Jiahao Xu, Wei Shao, Lihui Chen and Lemao Liu</td>
<td>12028</td>
</tr>
<tr>
<td>Unlearn What You Want to Forget: Efficient Unlearning for LLMs</td>
<td>Jiaao Chen and Diyi Yang</td>
<td>12041</td>
</tr>
</tbody>
</table>
Simplicity Level Estimate (SLE): A Learned Reference-Less Metric for Sentence Simplification
Liam Cripwell, Joël Legrand and Claire Gardent .................................................. 12053

Precedent-Enhanced Legal Judgment Prediction with LLM and Domain-Model Collaboration
Yiquan Wu, Siying Zhou, Yifei Liu, Weiming Lu, Xiaozhong Liu, Yating Zhang, Changlong Sun, Fei Wu and Kun Kuang ................................................................. 12060

FActScore: Fine-grained Atomic Evaluation of Factual Precision in Long Form Text Generation
Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen-tau Yih, Pang Wei Koh, Mohit Iyyer, Luke Zettlemoyer and Hannaneh Hajishirzi ................................................ 12076

Calc-X and Calcfomers: Empowering Arithmetical Chain-of-Thought through Interaction with Symbolic Systems
Marek Kdalčík, Michal Štefánik, Ondrej Sotolar and Vlastimil Martinek ............... 12101

CoF-CoT: Enhancing Large Language Models with Coarse-to-Fine Chain-of-Thought Prompting for Multi-domain NLU Tasks
Hoang H Nguyen, Ye Liu, Chenwei Zhang, Tao Zhang and Philip S. Yu .................... 12109

When Language Models Fall in Love: Animacy Processing in Transformer Language Models
Michael Hanna, Yonatan Belinkov and Sandro Pezzelle ........................................... 12120

Improving Unsupervised Relation Extraction by Augmenting Diverse Sentence Pairs
Qing Wang, Kang Zhou, Qiao Qiao, Yuepei Li and Qi Li ...................................... 12136

Paraphrase Types for Generation and Detection
Jan Philip Wahle, Bela Gipp and Terry Ruas ......................................................... 12148

Target-to-Source Augmentation for Aspect Sentiment Triplet Extraction
Yice Zhang, Yifan Yang, Meng Li, Bin Liang, Shiwei Chen and Ruifeng Xu .............. 12165

PAC-tuning: Fine-tuning Pre-trained Language Models with PAC-driven Perturbed Gradient Descent
Guangliang Liu, Zhiyu Xue, Xitong Zhang, Kristen Johnson and Rongrong Wang ......... 12178

Emergence of Abstract State Representations in Embodied Sequence Modeling
Tian Yun, Zilai Zeng, Kunal Handa, Ashish V Thapliyal, Bo Pang, Ellie Pavlick and Chen Sun 12190

Accelerating Toeplitz Neural Network with Constant-time Inference Complexity
Zhen Qin and Yiran Zhong .................................................................................... 12206

Dissecting Recall of Factual Associations in Auto-Regressive Language Models
Mor Geva, Jasmijn Bastings, Katja Filippova and Amir Globerson ............................ 12216

StereoMap: Quantifying the Awareness of Human-like Stereotypes in Large Language Models
Sullam Jeoung, Yubin Ge and Jana Diesner .............................................................. 12236

Select, Prompt, Filter: Distilling Large Language Models for Summarizing Conversations
Minh-Quang Pham, Sathish Reddy Indurthi, Shamil Chollampatt and Marco Turchi ...... 12257

Human Raters Cannot Distinguish English Translations from Original English Texts
Shira Wein ................................................................................................................ 12266

Impressions: Visual Semiotics and Aesthetic Impact Understanding
Julia Kruk, Caleb Ziems and Diyi Yang ................................................................. 12273
Identifying Statements Crucial for Awareness of Interpretive Nonsense to Prevent Communication Breakdowns
Tomoyuki Maekawa and Michita Imai .......................................................... 12550

Multilingual Large Language Models Are Not (Yet) Code-Switchers
Ruochen Zhang, Samuel Cahyawijaya, Jan Christian Blaise Cruz, Genta Indra Winata and Alham Fikri Aji .......................................................... 12567

Reinforced Target-driven Conversational Promotion
Huy Quang Dao, Lizi Liao, Dung D. Le and Yuxiang Nie .................................. 12583

Identification of Multimodal Stance Towards Frames of Communication
Maxwell Weinzierl and Sandra Harabagiu .......................................................... 12597

Unsupervised Sounding Pixel Learning
Yining Zhang, Yanli Ji and Yang Yang .......................................................... 12610

LM vs LM: Detecting Factual Errors via Cross Examination
Roi Cohen, May Hamri, Mor Geva and Amir Globerson .................................. 12621

Large Language Models: The Need for Nuance in Current Debates and a Pragmatic Perspective on Understanding
Bram van Dijk, Tom Kouwenhoven, Marco Spruit and Max Johannes van Duijn .................. 12641

PIEClass: Weakly-Supervised Text Classification with Prompting and Noise-Robust Iterative Ensemble Training
Yunyi Zhang, Minhao Jiang, Yu Meng, Yu Zhang and Jiawei Han .......................... 12655

MeaeQ: Mount Model Extraction Attacks with Efficient Queries
Chengwei Dai, Minxuan Lv, Kun Li and Wei Zhou ........................................ 12671

The CoT Collection: Improving Zero-shot and Few-shot Learning of Language Models via Chain-of-Thought Fine-Tuning
Seungone Kim, Se June Joo, Doyoung Kim, Joel Jang, Seonghyeon Ye, Jamin Shin and Minjoon Seo .......................................................... 12685

Explaining Interactions Between Text Spans
Sagnik Ray Choudhury, Pepa Atanasova and Isabelle Augenstein .......................... 12709

Predictive Chemistry Augmented with Text Retrieval
Yujie Qian, Zhening Li, Zhengkai Tu, Connor W. Coley and Regina Barzilay .................. 12731

System Combination via Quality Estimation for Grammatical Error Correction
Muhammad Reza Qorib and Hwee Tou Ng .......................................................... 12746

Rethinking Negative Pairs in Code Search
Haochen Li, Xin Zhou, Anh Tuan Luu and Chunyan Miao .................................. 12760

Question Answering as Programming for Solving Time-Sensitive Questions
Xinyu Zhu, Cheng Yang, Bei Chen, Siheng Li, Jian-Guang Lou and Yujiu Yang .................. 12775

Joint Geometrical and Statistical Domain Adaptation for Cross-domain Code Vulnerability Detection
Qianjin Du, Shiji Zhou, Xiaohui Kuang, Gang Zhao and Jidong Zhai ......................... 12791

Revisiting Sparse Retrieval for Few-shot Entity Linking
Yulin Chen, Zhenran Xu, Baotian Hu and Min Zhang ........................................ 12801
Controlling Pre-trained Language Models for Grade-Specific Text Simplification
Sweta Agrawal and Marine Carpuat ................................. 12807

CLEVR-Implicit: A Diagnostic Dataset for Implicit Reasoning in Referring Expression Comprehension
Jingwei Zhang, Xin Wu and Yi Cai .................................. 12820

“Are Your Explanations Reliable?” Investigating the Stability of LIME in Explaining Text Classifiers by Marrying XAI and Adversarial Attack
Christopher Burger, Lingwei Chen and Thai Le .......................... 12831

CQE: A Comprehensive Quantity Extractor
Satya Almasian, Vivian Kazakova, Philipp Göldner and Michael Gertz .................. 12845

Context Compression for Auto-regressive Transformers with Sentinel Tokens
Siyu Ren, Qi Jia and Kenny Q. Zhu ..................................... 12860

A Unified View of Evaluation Metrics for Structured Prediction
Yunmo Chen, William Gantt, Tongfei Chen, Aaron Steven White and Benjamin Van Durme 12868

A Deeper (Autoregressive) Approach to Non-Convergent Discourse Parsing
Oren Tsur and Yoav Tulpan ............................................. 12883

We are Who We Cite: Bridges of Influence Between Natural Language Processing and Other Academic Fields
Jan Philip Wahle, Terry Ruas, Mohamed Abdalla, Bela Gipp and Saif M. Mohammad .... 12896

Ties Matter: Meta-Evaluating Modern Metrics with Pairwise Accuracy and Tie Calibration
Daniel Deutsch, George Foster and Markus Freitag ........................ 12914

SODA: Million-scale Dialogue Distillation with Social Commonsense Contextualization
Hyunwoo Kim, Jack Hessel, Liwei Jiang, Peter West, Ximing Lu, Youngjiae Yu, Pei Zhou, Ronan Le Bras, Malihe Alikhani, Gunhee Kim, Maarten Sap and Yejin Choi .................. 12930

Multi-view Contrastive Learning for Entity Typing over Knowledge Graphs
Zhiwei Hu, Victor Gutierrez Basulto, Zhiliang Xiang, Ru Li and Jeff Z. Pan ............... 12950

MailEx: Email Event and Argument Extraction
Saurabh Srivastava, Gaurav Singh, Shou Matsumoto, Ali K Raz, Paulo Costa, Joshua Campbell Poore and Ziyu Yao .................................................... 12964

Optimized Tokenization for Transcribed Error Correction
Tomer Wullach and Shlomo Chazan .................................. 12988

Beware of Model Collapse! Fast and Stable Test-time Adaptation for Robust Question Answering
Yi Su, Yixin Ji, Juntao Li, Hai Ye and Min Zhang .......................... 12998

Generative Adversarial Training with Perturbed Token Detection for Model Robustness
Jiahao Zhao and Wenji Mao ............................................ 13012

Multi-Task Knowledge Distillation with Embedding Constraints for Scholarly Keyphrase Boundary Classification
Seo Yeon Park and Cornelia Caragea .................................. 13026

Set Learning for Generative Information Extraction
Jiangnan Li, Yice Zhang, Bin Liang, Kam-Fai Wong and Rui Feng Xu ...................... 13043

Large Language Models and Multimodal Retrieval for Visual Word Sense Disambiguation
Anastasia Kritharoula, Maria Lymperaiou and Giorgos Stamou .......................... 13053

lxxxvi
Yangyi Chen, Xingyao Wang, Manling Li, Derek Hoiem and Heng Ji 13342

LLMLingua: Compressing Prompts for Accelerated Inference of Large Language Models
Huiqiang Jiang, Qianhui Wu, Chin-Yew Lin, Yuqing Yang and Lili Qiu 13358

EXPLAIN, EDIT, GENERATE: Rationale-Sensitive Counterfactual Data Augmentation for Multi-hop Fact Verification
Yingjie Zhu, Jiasheng Si, Yibo Zhao, Haiyang Zhu, Deyu Zhou and Yulan He 13377

An Exploration of Left-Corner Transformations
Andreas Opedal, Eleftheria Tsipidi, Tiago Pimentel, Ryan Cotterell and Tim Vieira 13393

Characterizing and Verifying Scientific Claims: Qualitative Causal Structure is All You Need
Jinxuan Wu, Wenhan Chao, Xian Zhou and Zhunchen Luo 13428

FOCUS: Effective Embedding Initialization for Monolingual Specialization of Multilingual Models
Konstantin Dobler and Gerard de Melo 13440

ByteSized32: A Corpus and Challenge Task for Generating Task-Specific World Models Expressed as Text Games
Ruoyao Wang, Graham Todd, Xingdi Yuan, Zhang Xiao, Marc-Alexandre Côté and Peter Jansen 13455

Skill-Based Few-Shot Selection for In-Context Learning
Shengnan An, Bo Zhou, Zeqi Lin, Qiang Fu, Bei Chen, Nanning Zheng, Weizhu Chen and Jian-Guang Lou 13472

MaNtLE: Model-agnostic Natural Language Explainer
Rakesh R Menon, Kerem Zaman and Shashank Srivastava 13493

PTP: Boosting Stability and Performance of Prompt Tuning with Perturbation-Based Regularizer
Lichang Chen, Jiuhai Chen, Heng Huang and Minhao Cheng 13512

Ling-CL: Understanding NLP Models through Linguistic Curricula
Mohamed Elgaar and Hadi Amiri 13526

Towards Unsupervised Recognition of Token-level Semantic Differences in Related Documents
Jannis Vamvas and Rico Sennrich 13543

Towards a Better Understanding of Variations in Zero-Shot Neural Machine Translation Performance
Shaomu Tan and Christof Monz 13553

SEER : A Knapsack approach to Exemplar Selection for In-Context HybridQA
Jonathan Tonglet, Manon Reusens, Philipp Borchert and Bart Baesens 13569

Conversation Chronicles: Towards Diverse Temporal and Relational Dynamics in Multi-Session Conversations
Jihyoung Jang, Minseong Boo and Hyounghun Kim 13584

DueT: Image-Text Contrastive Transfer Learning with Dual-adapter Tuning
Taku Hasegawa, Kyosuke Nishida, Koki Maeda and Kuniko Saito 13607

Towards a Unified Conversational Recommendation System: Multi-task Learning via Contextualized Knowledge Distillation
Yeongseo Jung, Eunseo Jung and Lei Chen 13625
ClusterLLM: Large Language Models as a Guide for Text Clustering
Yuwei Zhang, Zihan Wang and Jingbo Shang ............................................................... 13903

Shuyan Zhou, Uri Alon, Sumit Agarwal and Graham Neubig .................................. 13921

Learn and Consolidate: Continual Adaptation for Zero-Shot and Multilingual Neural Machine Translation
Kaiyu Huang, Peng Li, Junpeng Liu, Maosong Sun and Yang Liu ............................... 13938

e-THERAPIST: I suggest you to cultivate a mindset of positivity and nurture uplifting thoughts
Kshitij Mishra, Priyanshu Priya, Manisha Burja and Asif Ekbal .................................. 13952

AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages

Quantifying Character Similarity with Vision Transformers
Xinmei Yang, Abhishek Arora, Shao-Yu Jheng and Melissa Dell ................................. 13982

Syllogistic Reasoning for Legal Judgment Analysis
Wentao Deng, Jiahuan Pei, Keyi Kong, Zhe Chen, Furu Wei, Yujun Li, Zhaochun Ren, Zhumin Chen and Pengjie Ren ................................................................. 13997

Improving Transformer-based Program Repair Model through False Behavior Diagnosis
Youngkyoung Kim, Misoo Kim and Eunseok Lee ....................................................... 14010

SUT: Active Defects Probing for Transcompiler Models
Mengnan Qi, Yufan Huang, Maoquan Wang, Yongqiang Yao, Zihan Liu, Bin Gu, Colin Clement and Neel Sundaresan ................................................................. 14024

KCTS: Knowledge-Constrained Tree Search Decoding with Token-Level Hallucination Detection
Sehyun Choi, Tianqing Fang, Zhaowei Wang and Yangqiu Song ................................. 14035

CRUSH4SQL: Collective Retrieval Using Schema Hallucination For Text2SQL
Mayank Kothyari, Dhruva Dhingra, Sunita Sarawagi and Soumen Chakrabarti ............... 14054

This Reads Like That: Deep Learning for Interpretable Natural Language Processing
Claudio Fanconi, Moritz Vandenhirtz, Severin Husmann and Julia E Vogt ................... 14067

Incorporating Structured Representations into Pretrained Vision & Language Models Using Scene Graphs
Roei Herzig, Alon Mendelson, Leonid Karlinsky, Assaf Arbelle, Rogerio Feres, Trevor Darrell and Amir Globerson ................................................................. 14077

TLM: Token-Level Masking for Transformers
Yangjun Wu, Kebin Fang, Dongxiang Zhang, Han Wang, Hao Zhang and Gang Chen .... 14099

Addressing NER Annotation Noises with Uncertainty-Guided Tree-Structured CRFs
Jian Liu, Weichang Liu, Yufeng Chen, Jinan Xu and Zhe Zhao .................................. 14112

Hi Guys or Hi Folks? Benchmarking Gender-Neutral Machine Translation with the GeNTE Corpus
Andrea Piergentili, Beatrice Savoldi, Dennis Fucci, Matteo Negri and Luisa Bentivogli ... 14124
Multilingual Holistic Bias: Extending Descriptors and Patterns to Unveil Demographic Biases in Languages at Scale

GlobalBench: A Benchmark for Global Progress in Natural Language Processing
Yueqi Song, Simran Khanuja, Pengfei Liu, Fahim Faisal, Alissa Ostapenko, Genta Indra Winata, Alham Fikri Aji, Samuel Cahyawijaya, Yulia Tsvetkov, Antonios Anastasopoulos and Graham Neubig 14157

DetGPT: Detect What You Need via Reasoning
Renjie Pi, Jiahui Gao, Shizhe Diao, Rui Pan, Hanze Dong, Jipeng Zhang, Lewei Yao, Jianhua Han, Hang Xu, Lingpeng Kong and Tong Zhang .......................... 14172

Language Models with Rationality
Nora Kassner, Oyvind Tafjord, Ashish Sabharwal, Kyle Richardson, Hinrich Schuetze and Peter Clark .......................................................... 14190

Self-Improvement of Non-autoregressive Model via Sequence-Level Distillation
Yusheng Liao, Shuyang Jiang, Yiqi Li, Yu Wang and Yanfeng Wang .......................... 14202

Mitigating Temporal Misalignment by Discarding Outdated Facts
Michael J.Q. Zhang and Eunsol Choi .......................................................... 14213

Open-world Semi-supervised Generalized Relation Discovery Aligned in a Real-world Setting
William P Hogan, Jiachen Li and Jingbo Shang .......................... 14227

IEKG: A Commonsense Knowledge Graph for Idiomatic Expressions
Ziheng Zeng, Kellen Tan Cheng, Srihari Venkat Nanniyur, Jianing Zhou and Suma Bhat . 14243

Bias Neutralization in Non-Parallel Texts: A Cyclic Approach with Auxiliary Guidance
Karthic Madanagopal and James Caverlee .......................................................... 14265

Fighting Fire with Fire: The Dual Role of LLMs in Crafting and Detecting Elusive Disinformation
Jason S Lucas, Adaku Uchendu, Michiharu Yamashita, Jooyoung Lee, Shaurya Rohatgi and Dong-won Lee .......................................................... 14279

SMoP: Towards Efficient and Effective Prompt Tuning with Sparse Mixture-of-Prompts
Joon-Young Choi, Junho Kim, Jun-Hyung Park, Wing-Lam Mok and SangKeun Lee 14306

BRAINTEASER: Lateral Thinking Puzzles for Large Language Models
Yifan Jiang, Filip Ilievski, Kaixin Ma and Zhivar Sourati .......................................................... 14317

When are Lemons Purple? The Concept Association Bias of Vision-Language Models
Yingtian Tang, Yutaro Yamada, Yoyo Minzhi Zhang and Ilker Yildirim 14333

What Comes Next? Evaluating Uncertainty in Neural Text Generators Against Human Production Variability
Mario Giulianelli, Joris Baan, Wilker Aziz, Raquel Fernández and Barbara Plank 14349

Text Representation Distillation via Information Bottleneck Principle
Yanzhao Zhang, Dingkun Long, Zehan Li and Pengjun Xie .......................................................... 14372

Let GPT be a Math Tutor: Teaching Math Word Problem Solvers with Customized Exercise Generation
Zhenwen Liang, Wenhao Yu, Tanmay Rajpurohit, Peter Clark, Xiangliang Zhang and Ashwin Kalyan .......................................................... 14384
UniChart: A Universal Vision-language Pretrained Model for Chart Comprehension and Reasoning
Ahmed Masry, Parsa Kavehzadeh, Do Xuan Long, Enamul Hoque and Shafiq Joty ............ 14662

Merging Experts into One: Improving Computational Efficiency of Mixture of Experts
Shwai He, Run-Ze Fan, Liang Ding, Li Shen, Tianyi Zhou and Dacheng Tao ................. 14685

Distance-Based Propagation for Efficient Knowledge Graph Reasoning
Harry Shomer, Yao Ma, Juanhui Li, Bo Wu, Charu C. Aggarwal and Jiliang Tang .......... 14692

What to Read in a Contract? Party-Specific Summarization of Legal Obligations, Entitlements, and Prohibitions
Abhilasha Sancheti, Aparna Garimella, Balaji Vasan Srinivasan and Rachel Rudinger ...... 14708

Enhancing Computation Efficiency in Large Language Models through Weight and Activation Quantization
Janghwan Lee, Minsoo Kim, Seungcheol Baek, Seok Joong Hwang, Wonyong Sung and Jungwook Choi ................................................................. 14726

CP-BCS: Binary Code Summarization Guided by Control Flow Graph and Pseudo Code
Tong Ye, Lingfei Wu, Tengfei Ma, Xuhong Zhang, Yangkai Du, Peiyu Liu, Shouling Ji and Wenhai Wang ................................................................. 14740

Assessing Step-by-Step Reasoning against Lexical Negation: A Case Study on Syllogism
Mengyu Ye, Tatsuki Kuribayashi, Jun Suzuki, Goro Kobayashi and Hiroaki Funayama .... 14753

Chain-of-Thought Tuning: Masked Language Models can also Think Step By Step in Natural Language Understanding
Caoyun Fan, Jidong Tian, Yitian Li, Wenqing Chen, Hao He and Yaohui Jin ............... 14774

Large Language Models are Complex Table Parsers
Bowen Zhao, Changkai Ji, Yuejie Zhang, Wen He, Yingwen Wang, Qing Wang, Rui Feng and Xiaobo Zhang ................................................................. 14786

R2H: Building Multimodal Navigation Helpers that Respond to Help Requests
Yue Fan, Jing Gu, Kaizhi Zheng and Xin Eric Wang ............................................. 14803

Speech-enriched Memory for Inference-time Adaptation of ASR Models to Word Dictionaries
Ashish Mittal, Sunita Sarawagi, Preethi Jyothi, George Saon and Gakuto Kurata ....... 14820

Generative Table Pre-training Empowers Models for Tabular Prediction
Tianping Zhang, Shaowen Wang, Shuicheng Yan, Li Jian and Qian Liu ..................... 14836

Learning to Describe for Predicting Zero-shot Drug-Drug Interactions
Fangqi Zhu, Yongqi Zhang, Lei Chen, Bing Qin and Ruifeng Xu ......................... 14855

A Simple Baseline for Knowledge-Based Visual Question Answering
Alexandros Xenos, Themis Stafylakis, Ioannis Patras and Georgios Tzimiropoulos .... 14871

Unveiling the Essence of Poetry: Introducing a Comprehensive Dataset and Benchmark for Poem Summarization
Ridwan Mahbub, Ifrad Towhid Khan, Samiha Shafiq Anuva, Md Shihab Shahriar, Md Tahmid Rahman Laskar and Sabbir Ahmed ............................................. 14878

Privacy Implications of Retrieval-Based Language Models
Yangsibo Huang, Samyak Gupta, Zexuan Zhong, Kai Li and Danqi Chen ............... 14887
IMTLab: An Open-Source Platform for Building, Evaluating, and Diagnosing Interactive Machine Translation Systems
Xu Huang, Zhirui Zhang, Ruizhe Gao, Yichao Du, Lemao Liu, Guoping Huang, Shuming Shi, Jiajun Chen and Shujian Huang ................................................................. 14903

Is ChatGPT Good at Search? Investigating Large Language Models as Re-Ranking Agents
Weiwei Sun, Lingyong Yan, Xinyu Ma, Shuaiqiang Wang, Pengjie Ren, Zhumin Chen, Dawei Yin and Zhaochun Ren ................................................................. 14918

DiNeR: A Large Realistic Dataset for Evaluating Compositional Generalization
Chengang Hu, Xiao Liu and Yansong Feng ................................................................. 14938

Can Pre-trained Vision and Language Models Answer Visual Information-Seeking Questions?
Yang Chen, Hexiang Hu, Yi Luan, Haitian Sun, Soravit Changpinyo, Alan Ritter and Ming-Wei Chang ................................................................. 14948

EDeR: Towards Understanding Dependency Relations Between Events
Ruiqi Li, Patrik Haslum and Leyang Cui ................................................................. 14969

It Ain’t Over: A Multi-aspect Diverse Math Word Problem Dataset
Jiwoo Kim, Youngbin Kim, Ilwoong Baek, JinYeong Bak and Jongwuk Lee ................................................................. 14984

Dr ChatGPT tell me what I want to hear: How different prompts impact health answer correctness
Bevan Koopman and Guido Zuccon ................................................................. 15012

kNN-LM Does Not Improve Open-ended Text Generation
Shufan Wang, Yixiao Song, Andrew Drozdov, Aparna Garimella, Varun Manjunatha and Mohit Iyyer ................................................................. 15023

Towards A Unified View of Sparse Feed-Forward Network in Pretraining Large Language Model
Zeyu Liu, Tim Dettmers, Xi Victoria Lin, Veselin Stoyanov and Xian Li ................................................................. 15038

Exploring the Impact of Model Scaling on Parameter-Efficient Tuning
Yusheng Su, Chi-Min Chan, Jiali Cheng, Yujia Qin, Yankai Lin, Shengding Hu, Zonghan Yang, Ning Ding, Xingzhi Sun, Guotong Xie, Zhiyuan Liu and Maosong Sun ................................................................. 15062

STAIR: Learning Sparse Text and Image Representation in Grounded Tokens
Chen Chen, Bowen Zhang, Liangliang Cao, Jiguang Shen, Tom Gunter, Albin Madappally Jose, Alexander T Toshev, Yantao Zheng, Jonathon Shlens, Ruoming Pang and Yinfei Yang ................................................................. 15079

Crossing the Threshold: Idiomatic Machine Translation through Retrieval Augmentation and Loss Weighting
Emmy Liu, Aditi Chaudhary and Graham Neubig ................................................................. 15095

CoRec: An Easy Approach for Coordination Recognition
Qing Wang, Haojie Jia, Wenfei Song and Qi Li ................................................................. 15112

A linear time approximation of Wasserstein distance with word embedding selection
Sho Otao and Makoto Yamada ................................................................. 15121

Exchange-of-Thought: Enhancing Large Language Model Capabilities through Cross-Model Communication
Zhangyue Yin, Qiushi Sun, Cheng Chang, Qipeng Guo, Junqi Dai, Xuanjing Huang and Xipeng Qiu ................................................................. 15135

Conversation Understanding using Relational Temporal Graph Neural Networks with Auxiliary Cross-Modality Interaction
Cam Van Thi Nguyen, Tuan Anh Mai, Son Le The, Dang Hai Kieu and Duc-Trong Le .... 15154
CorefPrompt: Prompt-based Event Coreference Resolution by Measuring Event Type and Argument Compatibilities
Sheng Xu, Peifeng Li and Qiaoming Zhu .......................................................... 15440

Anaphor Assisted Document-Level Relation Extraction
Chonggang Lu, Richong Zhang, Kai Sun, Jaein Kim, Cunwang Zhang and Yongyi Mao . . 15453

FinEntity: Entity-level Sentiment Classification for Financial Texts
Yixuan Tang, Yi Yang, Allen H Huang, Andy Tam and Justin Z. Tang ......................... 15465

All Things Considered: Detecting Partisan Events from News Media with Cross-Article Comparison
Yujian Liu, Xinliang Frederick Zhang, Kaijian Zou, Ruihong Huang, Nicholas Beauchamp and Lu Wang ................................................................. 15472

Rationale-Enhanced Language Models are Better Continual Relation Learners
Weimin Xiong, Yifan Song, Peiyi Wang and Sujian Li ............................................. 15489

BanglaAbuseMeme: A Dataset for Bengali Abusive Meme Classification
Mithun Das and Animesh Mukherjee ............................................................... 15498

ScanDL: A Diffusion Model for Generating Synthetic Scanpaths on Texts
Lena Sophia Bolliger, David Robert Reich, Patrick Haller, Deborah Noemie Jakobi, Paul Prasse and Lena Ann Jäger ..................................................... 15513

From Values to Opinions: Predicting Human Behaviors and Stances Using Value-Injected Large Language Models
Dongjun Kang, Joosuk Park, Yohan Jo and JinYeong Bak .................................... 15539

Analyzing Film Adaptation through Narrative Alignment
Tanzir Pial, Shahreen Salim Aunti, Charruta Pethe, Allen Kim and Steven Skiena .......... 15560

Inverse Scaling Can Become U-Shaped
Jason Wei, Najoung Kim, Yi Tay and Quoc V Le ................................................. 15580

Nearest Neighbor Machine Translation is Meta-Optimizer on Output Projection Layer
Ruize Gao, Zhirui Zhang, Yichao Du, Lemao Liu and Rui Wang ............................. 15592

Variance Matters: Detecting Semantic Differences without Corpus/Word Alignment
Ryo Nagata, Hiroya Takamura, Naoki Otani and Yoshifumi Kawasaki ................. 15609

MolCA: Molecular Graph-Language Modeling with Cross-Modal Projector and Uni-Modal Adapter
Zhiyuan Liu, Sihang Li, Yanchen Luo, Hao Fei, Yixin Cao, Kenji Kawaguchi, Xiang Wang and Tat-Seng Chua .......................................................... 15623

A Training-Free Debiasing Framework with Counterfactual Reasoning for Conversational Emotion Detection
Geng Tu, Ran Jing, Bin Liang, Min Yang, Kam-Fai Wong and Ruifeng Xu .............. 15639

Self-ICL: Zero-Shot In-Context Learning with Self-Generated Demonstrations
Wei-Lin Chen, Cheng-Kuang Wu, Yun-Nung Chen and Hsin-Hsi Chen .................. 15651

Learning Knowledge-Enhanced Contextual Language Representations for Domain Natural Language Understanding
Taolin Zhang, Ruyao Xu, Chengyu Wang, Zhongjie Duan, Cen Chen, Minghui Qiu, Dawei Cheng, Xiaofeng He and Weining Qian ........................................ 15663

ScdNER: Span-Based Consistency-Aware Document-Level Named Entity Recognition
Ying Wei and Qi Li  ......................................................................................... 15677
MQuAKE: Assessing Knowledge Editing in Language Models via Multi-Hop Questions
Zexuan Zhong, Zhengxuan Wu, Christopher D Manning, Christopher Potts and Danqi Chen

Stance Detection on Social Media with Background Knowledge
Ang Li, Bin Liang, Jingqian Zhao, Bowen Zhang, Min Yang and Ruifeng Xu

Vision-Enhanced Semantic Entity Recognition in Document Images via Visually-Asymmetric Consistency Learning
Hao Wang, Xiahua Chen, Rui Wang and Chenhui Chu

NormDial: A Comparable Bilingual Synthetic Dialog Dataset for Modeling Social Norm Adherence and Violation
Oliver Li, Mallika Subramanian, Arkadiy Saakyan, Sky CH-Wang and Smaranda Muresan

ClimateBERT-NetZero: Detecting and Assessing Net Zero and Reduction Targets
Tobias Schimanski, Julia Bingler, Mathias Kraus, Camilla Hyslop and Markus Leippold

Leap-of-Thought: Accelerating Transformers via Dynamic Token Routing
Yecheon Kim, Junho Kim, Jun-Hyung Park, Mingyu Lee and SangKeun Lee

Reinforcement Replaces Supervision: Query focused Summarization using Deep Reinforcement Learning
Swaroop Nath, Pushpak Bhattacharyya and Harshad Khadilkar

Fair Text Classification with Wasserstein Independence
Thibaud Leteno, Antoine Gourru, Charlotte Laclau, Rémi Emonet and Christophe Gravier

TacoPrompt: A Collaborative Multi-Task Prompt Learning Method for Self-Supervised Taxonomy Completion
Hongyuan Xu, Ciyi Liu, Yuhang Niu, Yunong Chen, Xiangrui Cai, Yanlong Wen and Xiaojie Yuan

An Attribution Method for Siamese Encoders
Lucas Moeller, Dmitry Nikolaev and Sebastian Padó

Global Voices, Local Biases: Socio-Cultural Prejudices across Languages
Anjishnu Mukherjee, Chahat Raj, Ziwei Zhu and Antonios Anastasopoulos

Graph vs. Sequence: An Empirical Study on Knowledge Forms for Knowledge-Grounded Dialogue
Yizhe Yang, Heyan Huang, Yuhang Liu and Yang Gao

Are Compressed Language Models Less Subgroup Robust?
Leonidas Gee, Andrea Zugarini and Novi Quadrianto

Length Does Matter: Summary Length can Bias Summarization Metrics
Xiaobo Guo and Soroush Vosoughi

NL2TL: Transforming Natural Languages to Temporal Logics using Large Language Models
Yongchao Chen, Rujul Gandhi, Yang Zhang and Chuchu Fan

Reformulating NLP tasks to Capture Longitudinal Manifestation of Language Disorders in People with Dementia.
Dimitris Gkoumas, Matthew Purver and Maria Liakata

Elevating Code-mixed Text Handling through Auditory Information of Words
Mamta Mamta, Zishan Ahmad and Asif Ekbal
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predict and Use: Harnessing Predicted Gaze to Improve Multimodal Sarcasm Detection</td>
<td>Divyank Pratap Tiwari, Diptesh Kanojia, Anupama Ray, Apoorva Nunna and Pushpak Bhattacharyya</td>
<td>15933</td>
</tr>
<tr>
<td>ViT-TTS: Visual Text-to-Speech with Scalable Diffusion Transformer</td>
<td>Huadai Liu, Rongjie Huang, Xuan Lin, Wenqiang Xu, Maozong Zheng, Hong Chen, Jinzheng He and Zhou Zhao</td>
<td>15957</td>
</tr>
<tr>
<td>Consistency Analysis of ChatGPT</td>
<td>Myeongjun Erik Jang and Thomas Lukasiewicz</td>
<td>15970</td>
</tr>
<tr>
<td>Do Differences in Values Influence Disagreements in Online Discussions?</td>
<td>Michiel van der Meer, Piek Vossen, Catholijn M Jonker and Pradeep Kumar Murukannaiah</td>
<td>15986</td>
</tr>
<tr>
<td>A Digital Language Coherence Marker for Monitoring Dementia</td>
<td>Dimitris Gkoumas, Adam Tsakalidis and Maria Liakata</td>
<td>16021</td>
</tr>
<tr>
<td>Detecting Spoilers in Movie Reviews with External Movie Knowledge and User Networks</td>
<td>Heng Wang, Wenqian Zhang, Yuyang Bai, Zhaoxuan Tan, Shangbin Feng, Qinghua Zheng and Minnan Luo</td>
<td>16035</td>
</tr>
<tr>
<td>Joyful: Joint Modality Fusion and Graph Contrastive Learning for Multimodal Emotion Recognition</td>
<td>Dongyuan Li, Yusong Wang, Kotaro Funakoshi and Manabu Okumura</td>
<td>16051</td>
</tr>
<tr>
<td>HyperRank: Hyperbolic Ranking Model for Unsupervised Keyphrase Extraction</td>
<td>Mingyang Song, Huafeng Liu and Liping Jing</td>
<td>16070</td>
</tr>
<tr>
<td>Assessing the influence of attractor-verb distance on grammatical agreement in humans and language models</td>
<td>Christos Nikolaos Zacharopoulos, Théo Desbordes and Mathias Sablé-Meyer</td>
<td>16081</td>
</tr>
<tr>
<td>Federated Meta-Learning for Emotion and Sentiment Aware Multi-modal Complaint Identification</td>
<td>Apoorva Singh, Siddarth Chandrasekar, Sriparna Saha and Tanmay Sen</td>
<td>16091</td>
</tr>
<tr>
<td>Semantic Similarity Models for Depression Severity Estimation</td>
<td>Anxo Pérez, Neha Warikoo, Kexin Wang, Javier Parapar and Iryna Gurevych</td>
<td>16104</td>
</tr>
<tr>
<td>Hop, Union, Generate: Explainable Multi-hop Reasoning without Rationale Supervision</td>
<td>Wenting Zhao, Justin T Chiu, Claire Cardie and Alexander M Rush</td>
<td>16119</td>
</tr>
<tr>
<td>To Split or Not to Split: Composing Compounds in Contextual Vector Spaces</td>
<td>Christopher William Jenkins, Filip Miletic and Sabine Schulte im Walde</td>
<td>16131</td>
</tr>
<tr>
<td>ToolWriter: Question Specific Tool Synthesis for Tabular Data</td>
<td>Carlos Gemmell and Jeff Dalton</td>
<td>16137</td>
</tr>
<tr>
<td>Title</td>
<td>Authors</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>AnyTOD: A Programmable Task-Oriented Dialog System</td>
<td>Jeffrey Zhao, Yuan Cao, Raghav Gupta, Harrison Lee, Abhinav Rastogi, Mingqiu Wang, Hagen Soltau, Izhak Shafran and Yonghui Wu</td>
<td>16189</td>
</tr>
<tr>
<td>Can LMs Generalize to Future Data? An Empirical Analysis on Text Summarization</td>
<td>Chi Seng Cheang, Hou Pong Chan, Derek F. Wong, Xuebo Liu, Zhaocong Li, Yanming Sun, Shudong Liu and Lidia S. Chao</td>
<td>16205</td>
</tr>
<tr>
<td>Zero-Shot Multi-Label Topic Inference with Sentence Encoders and LLMs</td>
<td>Souvika Sarkar, Dongji Feng and Shubhra Kanti Karmaker Santu</td>
<td>16218</td>
</tr>
<tr>
<td>TaskDiff: A Similarity Metric for Task-Oriented Conversations</td>
<td>Ankita Bhaumik, Praveen Venkateswaran, Yara Rizk and Vatche Isahagian</td>
<td>16234</td>
</tr>
<tr>
<td>Not all Fake News is Written: A Dataset and Analysis of Misleading Video Headlines</td>
<td>Yoo Yeon Sung, Jordan Lee Boyd-Graber and Naeemul Hassan</td>
<td>16241</td>
</tr>
<tr>
<td>Learning From Free-Text Human Feedback – Collect New Datasets Or Extend Existing Ones?</td>
<td>Dominic Petrak, Nafise Sadat Moosavi, Ye Tian, Nikolai Rozanov and Iryna Gurevych</td>
<td>16259</td>
</tr>
<tr>
<td>Euphemistic Abuse – A New Dataset and Classification Experiments for Implicitly Abusive Language</td>
<td>Michael Wiegand, Jana Kampfmeier, Elisabeth Eder and Josef Ruppenhofer</td>
<td>16280</td>
</tr>
<tr>
<td>Exploring Distributional Shifts in Large Language Models for Code Analysis</td>
<td>Shushan Arakelyan, Rocktim Jyoti Das, Yi Mao and Xiang Ren</td>
<td>16298</td>
</tr>
<tr>
<td>ATHENA: Mathematical Reasoning with Thought Expansion</td>
<td>JB. Kim, Hazel Kim, Joonghyuk Hahn and Yo-Sub Han</td>
<td>16315</td>
</tr>
<tr>
<td>A Benchmark for Reasoning with Spatial Prepositions</td>
<td>Iulia Maria Comsa and Srini Narayanan</td>
<td>16328</td>
</tr>
<tr>
<td>TIMELINE: Exhaustive Annotation of Temporal Relations Supporting the Automatic Ordering of Events in News Articles</td>
<td>Sarah Alsayyahi and Riza Batista-Navarro</td>
<td>16336</td>
</tr>
<tr>
<td>Mitigating Over-Generation for Unsupervised Keyphrase Extraction with Heterogeneous Centrality Detection</td>
<td>Mingyang Song, Pengyu Xu, Yi Feng, Huafeng Liu and Liping Jing</td>
<td>16349</td>
</tr>
<tr>
<td>Towards Interpretable and Efficient Automatic Reference-Based Summarization Evaluation</td>
<td>Yixin Liu, Alexander Fabbri, Yilun Zhao, Pengfei Liu, Shafiq Joty, Chien-Sheng Wu, Caiming Xiong and Dragomir Radev</td>
<td>16360</td>
</tr>
<tr>
<td>PK-ICR: Persona-Knowledge Interactive Multi-Context Retrieval for Grounded Dialogue</td>
<td>Minsik Oh, Joosung Lee, Jiwei Li and Guoyin Wang</td>
<td>16383</td>
</tr>
<tr>
<td>More Than Spoken Words: Nonverbal Message Extraction and Generation</td>
<td>Dian Yu, Xiaoyang Wang, Wanshun Chen, Nan Du, Longyue Wang, Haitao Mi and Dong Yu</td>
<td>16396</td>
</tr>
<tr>
<td>Title</td>
<td>Authors</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Can language models learn analogical reasoning? Investigating training objectives and comparisons to human performance</td>
<td>Molly Petersen and Lonneke van der Plas</td>
<td>16414</td>
</tr>
<tr>
<td>FAME: Flexible, Scalable Analogy Mappings Engine</td>
<td>Shahar Jacob, Chen Shani and Dafna Shahaf</td>
<td>16426</td>
</tr>
<tr>
<td>A Self-training Framework for Automated Medical Report Generation</td>
<td>Siyuan Wang, Zheng Liu and Bo Peng</td>
<td>16443</td>
</tr>
<tr>
<td>A Picture is Worth a Thousand Words: Language Models Plan from Pixels</td>
<td>Anthony Zhe Liu, Lajanugen Logeswaran, Sungryull Sohn and Honglak Lee</td>
<td>16450</td>
</tr>
<tr>
<td>Interpreting and Exploiting Functional Specialization in Multi-Head Attention under Multi-task Learning</td>
<td>Chong Li, Shaonan Wang, Yunhao Zhang, Jiajun Zhang and Chengqing Zong</td>
<td>16460</td>
</tr>
<tr>
<td>Multilingual Previously Fact-Checked Claim Retrieval</td>
<td>Matúš Pikuliak, Ivan Srba, Robert Moro, Timo Hromadka, Timotej Smoleň, Martin Melišek, Ivan Vykopal, Jakub Simko, Juraj Podroužek and Maria Bielikova</td>
<td>16477</td>
</tr>
<tr>
<td>ALCAP: Alignment-Augmented Music Captioner</td>
<td>Zihao He, Weituo Hao, Wei-Tsung Lu, Changyou Chen, Kristina Lerman and Xuchen Song</td>
<td>16501</td>
</tr>
<tr>
<td>Do Transformers Parse while Predicting the Masked Word?</td>
<td>Haoyu Zhao, Abhishek Panigrahi, Rong Ge and Sanjeev Arora</td>
<td>16513</td>
</tr>
<tr>
<td>Composable Text Controls in Latent Space with ODEs</td>
<td>Guangyi Liu, Zeyu Feng, Yuan Gao, Zichao Yang, Xiaodan Liang, Junwei Bao, Xiaodong He, Shuguang Cui, Zhen Li and Zhiting Hu</td>
<td>16543</td>
</tr>
<tr>
<td>P5: Plug-and-Play Persona Prompting for Personalized Response Selection</td>
<td>Joosung Lee, Minsik Oh and Donghun Lee</td>
<td>16571</td>
</tr>
<tr>
<td>Reader: Model-based language-instructed reinforcement learning</td>
<td>Nicola Dainese, Pekka Marttinen and Alexander Ilin</td>
<td>16583</td>
</tr>
<tr>
<td>Adapting Offline Speech Translation Models for Streaming with Future-Aware Distillation and Inference</td>
<td>Biao Fu, Minpeng Liao, Kai Fan, Zhongqiang Huang, Boxing Chen, Yidong Chen and Xiaodong Shi</td>
<td>16600</td>
</tr>
<tr>
<td>Relation-aware Ensemble Learning for Knowledge Graph Embedding</td>
<td>Ling Yue, Yongqi Zhang, Quanming Yao, Yong Li, Xian Wu, Ziheng Zhang, Zhenxi Lin and Yefeng Zheng</td>
<td>16620</td>
</tr>
<tr>
<td>Document-Level Machine Translation with Large Language Models</td>
<td>Longyue Wang, Chenyang Lyu, Tianbo Ji, Zhirui Zhang, Dian Yu, Shuming Shi and Zhaopeng Tu</td>
<td>16646</td>
</tr>
<tr>
<td>Multilingual Simplification of Medical Texts</td>
<td>Sebastian Antony Joseph, Kathryn Kazanas, Keziah Reina, Vishnesh J Ramanathan, Wei Xu, Byron C Wallace and Junyi Jessy Li</td>
<td>16662</td>
</tr>
<tr>
<td>Title</td>
<td>Authors</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>When Reviewers Lock Horns: Finding Disagreements in Scientific Peer Reviews</td>
<td>Sandeep Kumar, Tirthankar Ghosal and Asif Ekbal</td>
<td>16693</td>
</tr>
<tr>
<td>Argue with Me Tensely: Towards Sentence-Level Counter-Argument Generation</td>
<td>Jiayu Lin, Rong Ye, Meng Han, Qi Zhang, Ruofei Lai, Xinyu Zhang, Zhao Cao, Xuanjing Huang and Zhongyu Wei</td>
<td>16705</td>
</tr>
<tr>
<td>NLI4CT: Multi-Evidence Natural Language Inference for Clinical Trial Reports</td>
<td>Mael Jullien, Marco Valentino, Hannah Ruth Frost, Paul O’Regan, Dónal Landers and André Freitas</td>
<td>16745</td>
</tr>
<tr>
<td>Addressing Linguistic Bias through a Contrastive Analysis of Academic Writing in the NLP Domain</td>
<td>Robert Ridley, Zhen Wu, Jianbing Zhang, Shujian Huang and Xinyu Dai</td>
<td>16765</td>
</tr>
<tr>
<td>RobustGEC: Robust Grammatical Error Correction Against Subtle Context Perturbation</td>
<td>Yue Zhang, Leyang Cui, Enbo Zhao, Wei Bi and Shuming Shi</td>
<td>16780</td>
</tr>
<tr>
<td>Detecting Propaganda Techniques in Code-Switched Social Media Text</td>
<td>Muhammad Umar Salman, Asif Hanif, Shady Shehata and Preslav Nakov</td>
<td>16794</td>
</tr>
<tr>
<td>Speech Recognition and Meaning Interpretation: Towards Disambiguation of Structurally Ambiguous Spoken Utterances in Indonesian</td>
<td>Ruhiyah Faradishi Widiaputri, Ayu Purwarianti, Dessi Puji Lestari, Kurniawati Azizah, Dipta Tanaya and Sakriani Sakti</td>
<td>16813</td>
</tr>
<tr>
<td>Target-Agnostic Gender-Aware Contrastive Learning for Mitigating Bias in Multilingual Machine Translation</td>
<td>Minwoo Lee, Hyukhun Koh, Kang-il Lee, Dongdong Zhang, Minsung Kim and Kyomin Jung</td>
<td>16825</td>
</tr>
<tr>
<td>Code-Switching Metrics Using Intonation Units</td>
<td>Rebecca Pattichis, Dora LaCasse, Sonya Mitrovich Trawick and Rena Torres Cacoullos</td>
<td>16840</td>
</tr>
</tbody>
</table>