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

**Diamond and Welcome Event**

Grammarly

**Diamond**

LivePerson

**Platinum and D&I Ally**

Amazon Science

**Platinum**

Bloomberg Engineering
Silver

Bronze
©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

Message from the General Chair

Welcome to the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL). This is the flagship European conference dedicated to European and international researchers, covering a broad spectrum of research areas of Computational Linguistics and Natural Language Processing.

Organizing a scientific conference of the prestige and size of EACL is always a great honor associated with several challenges. Our team had to tackle unusual complexities: this conference was one of the first scheduled to be in person after the long period of online conferences forced by COVID pandemic. The bidding process for a location, which typically takes place several years before the actual start of the conference, is mainly driven by the aim of expanding and involving the science community of all European countries: EACL selected Kyiv, Ukraine, as the physical location. As you all know, in February 2022, an unpredictable and dramatic event happened, the war between Russian and Ukraine, which made the organization in Kyiv impossible.

Considering the importance of physical interaction among researchers, especially after the restrictions imposed by the COVID pandemic, we worked hard with the EACL and ACL boards to find an alternative location, able to delight our attendees. Our team achieved this seemingly impossible goal of organizing a conference in a new location a few months before its start: we selected Dubrovnik, Croatia, while preserving the original aim of strengthening the connection with the Ukrainian community. In this respect, the Ukraine local committee will feature a dedicated panel session, “Low-resource languages in NLP products”, and a workshop to highlight work on Ukrainian language technologies. Following the latest conference, EACL 2023 will be “hybrid,” serving both virtual and in-person participants. As our official local chairs are not from the physical location, we needed a local team from Croatia for helping with the logistics. As a result, the main unexpected novelty of EACL 2023 is to have two local organizing committees from two different European countries.

In the remainder of this preface, I would like to thank EACL contributors chronologically with respect to my work timeline for EACL: Roberto Basili and Shuly Wintner, the new and former Presidents of ACL, along with the EACL board – thanks for having trusted me to manage the organization of the conference in rather complicated times. I started to be confident that we would have done a good job after Isabelle Augenstein and Andreas Vlachos accepted the role of PC Chairs. They have performed amazing work, creating an outstanding program, and also helping me in recruiting our fantastic organization team. A special thank is due to Preslav Nakov (EACL officer) for his support: thanks to his action, the proactiveness of David Yarowsky, and the fairless effort of Jennifer Rachford (our new secretary of the ACL business office), we successfully implemented the apparently unrealistic idea of switching from the already planned online conference to a hybrid setting with a physical location in Dubrovnik. Regarding the online side of our hybrid conference, we partnered with Underline (Sol Rosenberg, Damira Mrsic and Luka Simic), who also gave us support for managing the entire conference. While finalizing the location, we started to activate the different sections of the conference, for which my acknowledgements are again in chronological order:

- Ukraine Local Committee, Viktoria Kolomiets, Mariana Romanyshyn, Oleksii Molchanovskyi, Oles Dobosevych, was instrumental in preserving our initial goal of connecting the Ukraine research community, organizing a panel and a workshop.
- The website chairs, Pepa Atanasova and Julius Cheng, started immediately to design our website, even when almost no information was available.
- The workshop chairs, Zeerak Talat and Antonio Toral, selected our conferences and led the selection of workshops for the joint ACL call.
• The tutorial chairs, Sameer Pradhan and Fabio Massimo Zanzotto, together with the ACL chairs, took care of the tutorial selection for the ACL related conferences.

• The demonstration chairs, Danilo Croce and Luca Soldaini, created a parallel conference program to select exciting demos.

• The Publicity Chairs, Laura Biester, Leshem Choshen and Joel Tetrault, have been our interface with the science community through social media platforms.

• The Publication Chairs, Carolina Scarlton and Ryan Cotterell, produced high-quality proceedings, thanks to their competence and experience.

• The diversity and inclusion chairs, Sara Tonelli, Elena Cabrio, Verena Rieser, Spandana Gella, took care of DI and performed an amazing job, also working on hundreds applications.

• The Local Organising Committee of Croatia, Marko Tadić, Krešimir Šojat, and Daša Farkaš, gave essential help for the logistics, Visa, and student volunteers.

• Student Research Workshop Chairs, Matthias Lindemann, Alban Petit, and Elisa Bassignana, along with their faculty advisors Valerio Basile and Natalie Schluter, helped in setting the bases for forming great NLP researchers of the future.

• Our entire program committee, Senior Area Chairs, Area Chairs, reviewers, and best paper committee, was essential for obtaining our high-quality scientific program.

• The ACL’s sponsorship director Chris Callison-Burch took care of our sponsorships.

• The student volunteers, as usual, are essential for a successful conference execution.

• Priscilla Rasmussen, our former ACL business office secretary, continued to provide us with useful advice.

Finally, I would like to thank our sponsors for helping us to fund scholarships and DI initiatives.

Alessandro Moschitti
Amazon Alexa AI, Los Angeles, USA
EACL 2023 General Chair
ACL Statement on the Ukraine situation

March 11, 2022

The Association for Computational Linguistics (ACL) condemns in the strongest possible terms the actions of the Russian Federation government in invading the sovereign state of Ukraine and engaging in war against the Ukrainian people. We stand together with Ukrainian NLP colleagues, the Ukrainian people, Russian NLP colleagues and Russian people who condemn the actions of the Russian Federation government, and all those around the world who have been impacted by the invasion.

As a small token of our solidarity with the Ukrainian people, the ACL has decided to temporarily sever its ties with Russia-based organizations, while at the same time allowing Russian scientists to remain part of the ACL community. In practice, this means that the ACL will refrain from accepting any sponsorship or allowing any exhibits from Russian-headquartered entities at ACL-run events. Russian scholars are still welcome to participate in ACL events and publish at ACL venues.

The ACL is committed to peace and condemns any form of violence and harassment. We are also committed to peaceful co-operation, mutual understanding, and tolerance across borders. NLP scholars from both Ukraine and Russia are welcome to get in touch with the ACL with any concerns.

Tim Baldwin, on behalf of the ACL Executive
Message from the Program Chairs

Welcome to the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL). After the last edition in 2021 having been held fully online due to the COVID pandemic, EACL 2023 is being held in “hybrid” mode this year, serving both virtual and in-person participants in Dubrovnik, Croatia. While the original plan was to hold the conference in Kyiv (which was the plan originally for EACL 2021), the ongoing war made the organisation in Ukraine impossible. In order to ensure that the original aim of strengthening the connections with the Ukrainian community is still served, our program features a dedicated session and a workshop to highlight work on Ukrainian language technologies.

Submission and Acceptance

EACL 2023 accepted direct submissions, as well as submissions via ARR. For direct submissions, abstracts were needed to be registered one week prior to the submission date.

In total, EACL 2023 received 1550 submissions, the largest number to date, with the 2021 edition having received 1400 submissions. Out of those, 1045 were long and 505 were short paper submissions. 81 were ARR papers that were committed to EACL. 249 submissions were withdrawn throughout the reviewing process, including before the full paper submission deadline. 55 papers were desk rejected for various reasons (missing the limitations section, anonymity policy, multiple submission policy, plagiarism or formatting violations).

By the time we as the programme chairs made acceptance decisions, 1166 submissions were still active in the system. We kept the acceptance rate in line with previous *ACL conferences, resulting in 281 papers accepted to the main conference (24.1%), and 201 papers accepted to the Findings of EACL (17.2%), with the remaining 58.7% being rejected. One paper accepted to the main conference and four papers accepted to Findings were subsequently withdrawn. Out of the final set of accepted main conference papers, we invited 178 to be presented orally, and all 281 papers accepted to the main conference to be presented during in-person sessions, as well as a plenary virtual poster session. The EACL 2023 program also features six papers from the Transactions of the Association for Computational Linguistics (TACL) journal, and one from the Computational Linguistics (CL) journal.

Limitations Section

Following EMNLP 2022, we required that each submitted paper must include an explicitly named Limitations section, discussing the limitations of the work. This was to counterbalance the practice of over-hyping the take-away messages of papers, and to encourage more rigorous and honest scientific practice. This discussion did not count towards the page limit, and we asked reviewers to not use the mentioned limitations as reasons to reject the paper, unless there was a really good reason to.

Areas

To ensure a smooth process, the submissions to EACL 2023 were divided into 21 areas. The areas mostly followed these of previous EACL, and more broadly *ACL conferences, reflecting the typical divisions in the field. We also had a special area for papers for which both SACs had a conflict of interest. Those papers were reviewed by the reviewers and ACs in their original areas, but the paper recommendations were made by a dedicated SAC, who was a senior member of the NLP community. The most popular areas with over 100 submissions were “Generation and Summarization”, “Language Resources and Evaluation”, and “Machine Learning in NLP”.

viii
Best Paper Awards

From the papers submitted to EACL 2023, we selected 25 papers accepted to the main conference as candidates for a Best Paper award, based on nominations by the reviewers. These papers were assessed by the Best Paper Award Committee, who also determined the types of paper awards, following the ACL Conference Awards Policy. The selected best papers and runner-ups will be announced in a dedicated plenary session for Best Paper Awards on 4 May 2023.

Programme Committee Structure and Reviewing

Similar to prior NLP conferences, we adopted the hierarchical program committee structure, where for each area we invited 1-2 Senior Area Chairs (SACs), who worked with a team of Area Chairs (ACs), and a larger team of reviewers. We relied on statistics from prior years to estimate how many SACs, ACs and reviewers would be needed and ended up with 43 SACs, 118 ACs and 1634 reviewers. For identifying ACs and reviewers, we used the reviewer lists from prior *ACL conferences, and also encouraged all EACL 2023 authors to serve as reviewers, using a mandatory form requesting further information on their ability to serve as ACs, reviewers or emergency reviewers, which authors had to fill in on Softconf when registering their abstracts. We passed this information on to SACs, who were responsible for recruiting ACs and reviewers.

Rather than making assignments using a matching algorithm, we asked ACs and reviewers to bid on registered abstracts within their areas, to achieve a better fit. We went with this solution as the number of papers per area was relatively small, and we wanted to avoid poor reviewing assignments as much as possible. We then made an initial paper assignment, in which we ensured that each paper would be reviewed by at least one reviewer who bidded “yes” for the submission, and by no reviewers who bidded “no” for the submission.

Afterwards, we asked the SACs to fine-tune the allocations, and ensure each paper had one AC and three reviewers assigned to it.

To ensure the review quality, we provided detailed guidelines about what reviewers should and shouldn’t do in a review, based on the EMNLP 2022 guidelines. We also asked reviewers to flag papers for potential ethical concerns.

For pre-reviewed ARR papers, we asked SACs to not rely mainly on the reviewer scores, but to make their recommendations based on the text of the reviews, meta-reviews and the papers themselves. For making acceptance decisions, we mostly followed SAC recommendations, though also taking into account the overall quality of papers submitted to the conference. Where recommendations seemed overly harsh or lenient given the reviewers’ scores, reviews, author responses, or discussions amongst reviewers, we engaged in a dialogue with the respective SACs to make the final decision about the papers in question.

Ethics Committee

We also formed an Ethics Committee (EC) dedicated to ethical issues. The ethics committee considered 21 papers that were flagged by the technical reviewing committee for ethical concerns. Out of these, 10 were conditionally accepted, meaning the ethics issues had to be addressed in the camera-ready version, to be verified by the EC prior to final acceptance, and the other 11 were accepted as is. The authors of all conditionally accepted papers submitted the camera-ready version and a short response that explained how they had made the changes requested by the EC. The EC double-checked these revised submissions and responses, and confirmed that the ethical concerns had been addressed. As a result, all conditionally accepted papers were accepted to the main conference or Findings.
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 EACL 2023, we decided to follow EMNLP 2022’s example and run a process which is separate from ARR, but also allows for ARR submissions. Specifically, authors could either submit papers to EACL 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 from the OpenReview system, according to a submission link that the author provided when committing their ARR submission to EACL. The ARR commitment deadline was set one month after the direct submission deadline since the ARR submissions already have their reviews and meta-recommendation. These ARR papers were then ranked by the SACs together with the direct submissions in the track, and based on the reviews and meta-reviews from ARR. Overall, EACL had 81 papers committed from ARR, of these 24 were accepted to the main conference and 20 were accepted to Findings of EACL.

Presentation Mode

We made the decision on which papers would be invited for oral poster presentations based on several factors: the relative rank of the paper according to SAC recommendation, whether the paper had been recommended for a best paper award by at least one reviewer, and for TACL and CL papers, the authors’ preference of presentation mode.

Keynotes and Panel

Another highlight of our program are the plenary sessions, for which we scheduled three talks, as well a panel:

- a keynote talk by Joyce Chai (University of Michigan) on “Language Use in Embodied AI!
- a keynote talk by Edward Greffenstette (Cohere AI and University College London) on “Going beyond the benefits of scale by reasoning about data”
- a keynote talk by Kevin Munger (Penn State University) on Chatbots for Good and Evil”
- a panel on “low-resource languages in NLP products” led by Mariana Romanysnhy with Viktoria Kolomiets (Grammarly), Mariana Romanysnhy (Grammarly), Oleksii Molchanovskyi (Ukrainian Catholic University) and Oles Dobosevych (Ukrainian Catholic University)

Thank Yous

EACL 2023 is the result of a collaborative effort and a supportive community, and we want to acknowledge the efforts of so many people with whom we worked directly and made significant efforts in putting together the programme for EACL 2023!

- Our General Chair, Alessandro Moschitti, who led the whole organising team, and helped with many of the decision processes;
- Our 43 Senior Area Chairs, who were instrumental in every aspect of the review process, from recruiting Area Chairs, correcting reviewer assignments, to making paper acceptances;
- Our 118 Area Chairs, who had the role of interacting with the reviewers, leading paper review discussions, and writing meta-reviews;
• The 1634 reviewers, who provided valuable feedback to the authors; The emergency reviewers, who provided their support at the last minute to ensure a timely reviewing process;

• Our Best Paper Selection Committee, who selected the best papers and the outstanding papers: Jonathan Kummerfeld (chair), Joakim Nivre, Bonnie Webber, Thamar Solorio and Hanna Hajishirzi;

• Our Ethics Committee, chaired by Zeerak Talat, for their hard work to ensure that all the accepted papers addressed the ethical issues appropriately, under a very tight schedule;

• Our amazing Publication Chairs, Carolina Scarton and Ryan Cotterell for compiling the proceedings in good time for the conference;

• Our Publicity Chairs, Laura Biester, Leshem Choshen and Joel Tetrault, for their work on managing the communications on social media platforms;

• Our website chairs, Pepa Atanasova and Julius Cheng for putting together the website for the conference and keeping it up to date;

• Damira Mrsic from Underline, for her support in developing the virtual conference platform;

• Jennifer Rachford, who has worked tirelessly online and on-site to ensure that EACL 2023 is a success.

We're looking forward to a great EACL 2023!

Isabelle Augenstein (University of Copenhagen, Denmark)
Andreas Vlachos (University of Cambridge, UK)
EACL 2023 Programme Committee Co-Chairs
Organizing Committee

General Chair
Alessandro Moschitti, Amazon Alexa

Program Chairs
Isabelle Augenstein, University of Copenhagen
Andreas Vlachos, University of Cambridge

Publications Chairs
Ryan Cotterell, ETH Zürich
Carolina Scarton, University of Sheffield

Workshop Chairs
Zeerak Talat, Simon Fraser University
Antonio Toral, University of Groningen

Tutorials Chairs
Sameer Pradhan, University of Pennsylvania
Fabio Massimo Zanzotto, University of Rome, “Tor Vergata”

Demonstrations Chairs
Danilo Croce, University of Rome, “Tor Vergata”
Luca Soldaini, Allen Institute for AI

Publicity Chairs
Joel Tetreault, Dataminr
Leshem Choshen, IBM AI research; Hebrew University of Jerusalem
Laura Biester, University of Michigan

Website Chairs
Pepa Atanasova, University of Copenhagen
Julius Cheng, University of Cambridge

Sponsorship Director
Chris Callison-Burch, University of Pennsylvania

Diversity and Inclusion Chairs
Elena Cabrio, Université Côte d’Azur, Inria, CNRS, I3S
Student Research Workshop Chairs
  Matthias Lindemann, University of Edinburgh
  Alban Petit, Université Paris-Saclay
  Elisa Bassignana, IT University of Copenhagen

Student Research Workshop Faculty Advisors
  Valerio Basile, University of Turin
  Natalie Schluter, IT University of Copenhagen; Apple

Local Organising Committee
  Marko Tadić, University of Zagreb
  Krešimir Šojat, University of Zagreb
  Daša Farkaš, University of Zagreb

Ukraine Local Committee
  Viktoria Kolomiets, Grammarly
  Mariana Romanyshyn, Grammarly
  Oleksii Molchanovskyi, Ukrainian Catholic University
  Oles Dobosevych, Ukrainian Catholic University
Program Committee

**Anaphora, Discourse and Pragmatics**
- Bonnie Webber, University of Edinburgh
- Michael Strube, Heidelberg Institute for Theoretical Studies

**Computational Social Science and Social Media**
- Maria Liakata, Queen Mary University of London
- Kalina Bontcheva, University of Sheffield

**Conflicts of Interests**
- Joakim Nivre, Research Institutes of Sweden

**Dialogue and Interactive Systems**
- Diarmuid Ó Séaghdha, Apple
- Matthew Purver, Queen Mary University of London

**Document analysis, Text Categorization and Topic Models**
- Nikolaos Aletras, University of Sheffield
- Ekaterina Shutova, University of Amsterdam

**Ethical and Sustainable NLP**
- Nafise Sadat Moosavi, Department of Computer Science, The University of Sheffield
- Yonatan Belinkov, Technion

**Ethics Review**
- Zeerak Talat, Simon Fraser University

**Generation and Summarization**
- Ondrej Dusek, Charles University
- Chenghua Lin, Department of Computer Science, University of Sheffield

**Information Extraction**
- Roberto Navigli, Sapienza University of Rome
- Mrinmaya Sachan, ETH Zurich

**Information Retrieval and Search**
- Bruno Martins, IST and INESC-ID
- Fabrizio Silvestri, Sapienza, University of Rome
Interpretability and Model Analysis

Dong Nguyen, Utrecht University
Roi Reichart, Technion - Israel Institute of Technology

Language Grounding and Multi-Modality

Grzegorz Chrupała, Tilburg University
Desmond Elliott, University of Copenhagen

Language Resources and Evaluation

Roman Klinger, University of Stuttgart
Omri Abend, The Hebrew University of Jerusalem

Linguistic Theories, Cognitive Modeling and Psycholinguistics

Barry Devereux, Queen’s University, Belfast
Natalie Schluter, IT University of Copenhagen

Machine Learning for NLP

James Henderson, Idiap Research Institute
Vlad Niculae, University of Amsterdam

Machine Translation

Wilker Aziz, University of Amsterdam
Rico Sennrich, University of Zurich

Multidisciplinary and other NLP Applications

Annie Priyadarshini Louis, Google Research UK
Yulan He, King’s College London

Multilinguality

Ivan Vulić, University of Cambridge
Alexander Fraser, Ludwig-Maximilians-Universität München

Phonology, Morphology, and Word Segmentation

Thierry Poibeau, LATTICE (CNRS and ENS/PSL)
François Yvon, ISIR CNRS and Sorbonne Université

Question Answering

Jonathan Berant, Tel Aviv University and AI2
Pontus Stenetorp, University College London
Semantics: lexical
Chris Biemann, Universität Hamburg
Mark Stevenson, University of Sheffield

Semantics: sentence level and other areas
Aliaksei Severyn, Google
Douwe Kiela, Hugging Face

Sentiment Analysis and Argument Mining
Veronique Hoste, LT3, Ghent University
Ivan Habernal, Technical University of Darmstadt

Tagging, Chunking, Syntax and Parsing
Marco Kuhlmann, Linköping University
Shay B. Cohen, University of Edinburgh

Area Chairs

Reviewers


Oliver Eberle, Sauleh Eetemadi, Steffen Eger, Annerose Eichel, Julian Eisen- schlos, Heba Elfardy, Micha Elsner, Saman Enayati, Aykut Erdem, Akiko Eriguchi, Katrin Erk, Ramy Eskander

Alex Fabbri, Marzieh Fadaee, Fahim Faisal, Neele Falk, Federico Fancellu, Qixiang Fang, Hossein Fani, Stefano Faralli, Oladimeji Farri, Nawshad Farruque, Manaal Faruqui, Mehrwish Fatima, Adam Faulkner, Pedro Faustini, Marc Feger, Nils Feldhus, Anna Feldman, Ghazi Felhi, Mariano Felice, Weixi Feng, Yue Feng, Manos Fergadiotis, Patrick Fernandes, Daniel Fernández-gonzález, Elisabetta Fersini, George Filandrianos, Elena Filatova, Mark Fishel, Lucie Flek, Michael Flor, Negar Foroutan Eghlidi, Jennifer Foster, Stella Frank, Jesse Freitas, Simona Frenna, Annemarie Friedrich, Lisheng Fu, Fumiyo Fukumoto, Kotaro Funakoshi


Le An Ha, Katharina Haemmerl, Gholamreza Haffari, Joonghyuk Hahn, Michael Hahn, Udo Ha-hn, Eva Hajicova, Dilek Hakkani-tur, Kishaloy Halder, Karina Halevy, Jiuzhou Han, Lifeng Han, Ting Han, Xudong Han, Yo-sub Han, Viktor Hangya, Sandra Harabagiu, Mareike Hartmann, Sadid A. Hasan, Sabit Hassan, Nabil Hathout, Amartya Hatua, Annette Hautli-Janisz, Adi Haviv, Yoshihiko Hayashi, Shirley Anugrah Hayati, T. J. Hazen, Rishi Hazra, Han He, Wanwei He, Wei He, Xiaoting He, Xuanli He, Xuehai He, Yun He, Behnam Hedayatnia, Kevin Heffernan, Benjamin Heinzloerling, Jindrich Helcl, William Held, Leonhard Hennig, Christian Herold, Jonathan Herzig, Gerhard Heyer, Derrick Higgins, Anthony Hills, Tatsuya Hiraoka, Vinh Thinh Ho, Cuong Hoang, Eben Holderness, Takeshi Homma, Ales Horák, Andrea Horbach, Sho Hoshino, Md Azam Hossain, Feng Hou, Yifan Hou, Yufang Hou, Shu-kai Hsieh, I-hung Hsu, Han Hu, Po Hu, Xinyu Hua, Chieh-yang Huang, Fei Huang, Hensch Huang, Jie Huang, Junbo Huang, Kuan-hao Huang, Quzhe Huang, Zhiqi Huang, Vojtech Hudecek, Pere-Lluis Huguet Cabot, Kai Hui, Chia-chien Hung, Julie Hunter

Nikolai Ilinykh, Dmitry Ilvovsky, Michimasa Inaba, Diana Inkpen, Koji Inoue, Hayate Iso, Takumi Ito, Maor Ivgi, Kenichi Iwatsuki, Vivek Iyer, Peter Izsak

Cassandra L. Jacobs, Sarthak Jain, Masoud Jalili Sabet, Sepehr Janghorbani, Adam Jatowt, Inigo Jauregi Unanue, Ganesh Jawahar, Harsh Jhamtani, Shaoxiong Ji, Yangfeng Ji, Chengyue Jiang, Junfeng Jiang, Longquan Jiang, Ming Jiang, Yuchen Eleanor Jiang, Ziyang Jiang, Baoyu Jing, Unso Jo, Richard Johansson, Aditya Joshi, Rishabh Joshi, Tahee Jung

Outstanding Reviewers

Keynote Talk: Going beyond the benefits of scale by reasoning about data

Edward Grefenstette
Cohere

Tuesday, May 2, 2023 – Time: 09:30 – 10:30 – Room: Elafiti 1, 2, 3 & 4

Abstract:
Transformer-based Large Language Models (LLMs) have taken NLP—and the world—by storm. This inflection point in our field marks a shift from focussing on domain-specific neural architecture design and the development of novel optimization techniques and objectives to a renewed focus on the scaling of model size and of the amount of data ingested during training. This paradigm shift yields surprising and delightful applications of LLMs, such as open-ended conversation, code understanding and synthesis, some degree of tool-use, and some zero-shot instruction-following capabilities. In this talk, I outline and lightly speculate on the mechanisms and properties which enable these diverse applications, and posit that the training regimen which enables these capabilities points to a further shift, namely one where we go from focussing on scale, to focussing on reasoning about what data to train on. I will briefly discuss recent advances in open-ended learning in Reinforcement Learning, and how some of the concepts at play in that work may inspire or directly apply to the development of novel ways of reasoning about data in supervised learning, in particular in areas pertaining to LLMs.

Bio:
Ed Grefenstette is the Head of Machine Learning at Cohere, a provider of cutting-edge NLP models that’s solving all kinds of language problems; including text summarization, composition, classification and more. In addition, Ed is an Honorary Professor at UCL. Ed’s previous industry experience comprises Facebook AI Research (FAIR), DeepMind, and Dark Blue Labs, where he was the CTO (acquired by Google in 2014). Prior to this, Ed worked at the University of Oxford’s Department of Computer Science, and was a Fulford Junior Research Fellow at Somerville College, whilst also lecturing students at Hertford College taking Oxford’s new computer science and philosophy course. Ed’s research interests span several topics, including natural language and generation, machine reasoning, open ended learning, and meta-learning.
Keynote Talk: Chatbots for Good and Evil

Kevin Munger
Penn State University

Wednesday, May 3, 2023 – Time: 15:45 – 16:45 – Room: Elafiti 1, 2, 3 & 4

Abstract:
The capacities of LLM-powered chatbots have been progressing on the order of months and have recently passed into mainstream public awareness and adoption. These tools have been used for a variety of scientific and policy interventions, but these advances call for a significant re-thinking of their place in society. Psychological research suggests that intentionality is a key factor in persuasion and social norm enforcement, and the proliferation of LLMs represents a significant shock to the intentionality contained in text and particularly in immediate, personalized chat. I argue that we are in a period of informational disequilibrium, where different actors have different levels of awareness of this technological shock. This period may thus represent a golden age for actors aiming to use these technologies at scale, for any number of normative ends; this includes social scientists and computational linguists. More broadly, I argue that the ethical frameworks for evaluating research practices using LLM-powered chatbots are insufficient to the scale of the current challenge. This is a potentially revolutionary technology that requires thinking in moral and political terms: given the power imbalances involved, it is of paramount importance that chatbots for good do not inadvertently become chatbots for evil.

Bio:
Kevin Munger is the Jeffrey L. Hyde and Sharon D. Hyde and Political Science Board of Visitors Early Career Professor of Political Science and Assistant Professor of Political Science and Social Data Analytics at Penn State University. Kevin’s research focuses on the implications of the internet and social media for the communication of political information. His specialty is the investigation of the economics of online media; current research models Clickbait Media and uses digital experiments to test the implications of these models on consumers of political information.
Keynote Talk: Language Use in Embodied AI

Joyce Chai
University of Michigan

Thursday, May 4, 2023 – Time: 14:15 – 15:15 – Room: Elafiti 1, 2, 3 & 4

Abstract:
With the emergence of a new generation of embodied AI agents, it becomes increasingly important to enable language communication between humans and agents. Language plays many important roles in embodied AI. In this talk, I will share some of the experiences in my lab that study the pragmatics of language, for example, in mediating perceptual differences, learning from language instructions, and planning for joint tasks. I will talk about how the embodied context shapes language use and influences computational models for language grounding to perception and action. I will show the importance of collaborative effort and theory of mind in language communication and how they affect common ground for situated tasks. I will discuss key challenges as well as new perspectives on these problems brought by recent advances in LLM and generative AI.

Bio:
Joyce Chai is a Professor in the Department of Electrical Engineering and Computer Science at the University of Michigan. Before joining UM in 2019, she was a Professor of Computer Science and Engineering at Michigan State University. She holds a Ph.D. in Computer Science from Duke University. Her research interests span from natural language processing and embodied AI to human-AI collaboration. She is fascinated by how experience with the world and how social pragmatics shape language learning and language use; and is excited about developing language technology that is sensorimotor grounded, pragmatically rich, and cognitively motivated. Her current work explores the intersection between language, perception, and action to enable situated communication with embodied agents. She served on the executive board of NAACL and as Program Co-Chair for multiple conferences – most recently ACL 2020. She is a recipient of the National Science Foundation Career Award and has received several paper awards with her students (e.g., the Best Long Paper Award at ACL 2010 and an Outstanding Paper Award at EMNLP 2021). She is a Fellow of ACL.
Panel: Low-resource languages in NLP products

Wednesday, May 3, 2023 – Time: 16:30 – 18:00 – Room: Elafiti 1, 2, 3 & 4

The panel discussion will bring together experts from industry and academia to share their experience building solutions for low-resource languages. We anticipate a lively discussion about the advantages and limitations of multilingual solutions and language-specific models, the challenges of evaluating models for low-resource languages, and the level of language awareness needed in the development process. In addition, the panelists will explore ways to increase the acceptance rate of papers that target low-resource languages at *ACL conferences. We hope that the panel discussion will increase the visibility of research for low-resource languages and emphasize its relevance.

Moderator: Mariana Romanyshyn, Grammarly
Mariana Romanyshyn is an Area Tech Lead for Computational Linguistics at Grammarly, Ukraine. She has professional experience in syntactic parsing, sentiment analysis, named entity recognition, fact extraction, and text anonymization. For the last eight years, Mariana has been working on error correction and text improvement algorithms at Grammarly. Mariana is an active speaker at AI conferences, co-organizer of the yearly Grammarly CompLing Summer School, co-organizer of the UNLP workshop, struggling reformer of Ukrainian university syllabuses, and active contributor of the Lang-uk group, focused on advancements in Ukrainian NLP.

Panelists:

Antonios Anastasopoulos, George Mason University
Antonios Anastasopoulos is an Assistant Professor in Computer Science at George Mason University. He received his PhD in Computer Science from the University of Notre Dame and then did a postdoc at Language Technologies Institute at Carnegie Mellon University. He also holds a BSc-MSc in Electrical and Computer Engineering from the National Technical University of Athens, Greece. His research is on natural language processing with a focus on multilinguality, low-resource settings, cross-lingual learning, and endangered languages, with the ultimate goal of building language technologies for under-served communities around the world. He is currently funded by the NSF, the NEH, the US DoD, Google, Amazon, and Meta.

Mona Diab, Meta
Mona Diab is the Lead Responsible AI Research Scientist with Meta. She is also a full Professor of Computer Science at the George Washington University (on leave) where she directs the CARE4Lang NLP Lab. Before joining Meta, she led the Lex Conversational AI project within Amazon AWS AI. Her current focus is on Responsible AI and how to operationalize it for NLP technologies. Her interests span
building robust technologies for low-resource scenarios with a special interest in Arabic technologies, (mis) information propagation, computational socio-pragmatics, computational psycholinguistics, NLG evaluation metrics, language modeling, and resource creation.

**Julia Makogon, Semantrum**

Julia Makogon is a Lead ML/NLP Engineer at Semantrum, a Ukrainian AI company that specializes in media analytics and reputation management. She studied Applied Mathematics at DSTU, Kamyanske, Ukraine, before pursuing a career in NLP. Julia developed multiple NLP applications for media monitoring, sentiment analysis, and legal document analysis for Ukrainian and other European languages. Her expertise lies in building industry solutions with limited resources. Julia serves at the Program Committee of the Ukrainian NLP workshop and is passionate about advancing solutions for the Ukrainian language.

**Ivan Vulić, University of Cambridge**

Ivan Vulić is a Principal Research Associate and a Royal Society University Research Fellow in the Language Technology Lab, University of Cambridge. He is also a Senior Scientist at PolyAI. He is a member of the Steering Committee of the Centre for Human Inspired Artificial Intelligence (CHIA) at Cambridge. Ivan holds a PhD in Computer Science from KU Leuven awarded summa cum laude. In 2021 he was awarded the annual Karen Spärck Jones Award from the British Computing Society for his research contributions to NLP and Information Retrieval. His core expertise is in representation learning, cross-lingual learning, conversational AI, human language understanding, distributional, lexical, multi-modal, and knowledge-enhanced semantics in monolingual and multilingual contexts, transfer learning for enabling cross-lingual NLP applications such as conversational AI in low-resource languages, and machine learning for (cross-lingual and multilingual) NLP. He has published numerous papers at top-tier NLP and Information Retrieval conferences and journals, and his research work also resulted in several best paper awards. He serves as an area chair and regularly reviews for all major NLP and Machine Learning conferences and journals. Ivan has given numerous invited talks at academia and industry and co-organised a number of NLP conferences and workshops.
# Table of Contents

**PiC: A Phrase-in-Context Dataset for Phrase Understanding and Semantic Search**  
Thang Pham, Seunghyun Yoon, Trung Bui and Anh Nguyen .................................. 1

**Enhancing Dialogue Summarization with Topic-Aware Global- and Local- Level Centrality**  
Xinnian Liang, Shuangzhi Wu, Chenhao Cui, Jiaqi Bai, Chao Bian and Zhoujun Li ........ 27

**Exploiting Summarization Data to Help Text Simplification**  
Renliang Sun, Zhixian Yang and Xiaojun Wan ......................................................... 39

**Shironaam: Bengali News Headline Generation using Auxiliary Information**  
Abu Ubaida Akash, Mir Tafseer Nayeem, Faisal Tareque Shohan and Tanvir Islam ........ 52

**PCC: Paraphrasing with Bottom-k Sampling and Cyclic Learning for Curriculum Data Augmentation**  
Hongyuan Lu and Wai Lam ....................................................................................... 68

**A Two-Sided Discussion of Preregistration of NLP Research**  
Anders Søgaard, Daniel Hershcovich and Miryam de Lhoneux ................................. 83

**WinoDict: Probing language models for in-context word acquisition**  

**Sentiment as an Ordinal Latent Variable**  
Niklas Stoehr, Ryan Cotterell and Aaron Schein ....................................................... 103

**Nationality Bias in Text Generation**  
Pranav Narayanan Venkit, Sanjana Gautam, Ruchi Panchanadikar, Ting-Hao Huang and Shomir Wilson ................................................................. 116

**Investigating data partitioning strategies for crosslinguistic low-resource ASR evaluation**  
Zoey Liu, Justin Spence and Emily Prud’hommeaux ................................................. 123

**Shortcomings of Question Answering Based Factuality Frameworks for Error Localization**  
Ryo Kamoi, Tanya Goyal and Greg Durrett .............................................................. 132

**Socratic Question Generation: A Novel Dataset, Models, and Evaluation**  
Beng Heng Ang, Sujatha Das Gollapalli and See-Kiong Ng ..................................... 147

**Do we need Label Regularization to Fine-tune Pre-trained Language Models?**  
Ivan Kobyzev, Aref Jafari, Mehdi Rezagholizadeh, Tianda Li, Alan Do-Omri, Peng Lu, Pascal Poupart and Ali Ghodsi ................................................................. 166

**COVID-VTS: Fact Extraction and Verification on Short Video Platforms**  
Fuxiao Liu, Yaser Yacoob and Abhinav Shrivastava .................................................. 178

**Multimodal Graph Transformer for Multimodal Question Answering**  
Xuehai He and Xin Wang ...................................................................................... 189

**Retrieval Enhanced Data Augmentation for Question Answering on Privacy Policies**  
Md Rizwan Parvez, Jianfeng Chi, Wasi Uddin Ahmad, Yuan Tian and Kai-Wei Chang ... 201

**FastKASSIM: A Fast Tree Kernel-Based Syntactic Similarity Metric**  
Maximillian Chen, Caitlyn Chen, Xiao Yu and Zhou Yu ........................................ 211

**Friend-training: Learning from Models of Different but Related Tasks**  
Mian Zhang, Lifeng Jin, Linfeng Song, Haitao Mi, Xiabing Zhou and Dong Yu .......... 232
Understanding Transformer Memorization Recall Through Idioms
Adi Haviv, Ido Cohen, Jacob Gidron, Roei Schuster, Yoav Goldberg and Mor Geva ........... 248

A Discerning Several Thousand Judgments: GPT-3 Rates the Article + Adjective + Numeral + Noun Construction
Kyle Mahowald ................................................................. 265

Triple-Hybrid Energy-based Model Makes Better Calibrated Natural Language Understanding Models
Haotian Xu and Yingying Zhang ............................................. 274

A weakly supervised textual entailment approach to zero-shot text classification
Marc Pàmies, Joan Llop, Francesco Multari, Nicolau Duran-Silva, César Parra-Rojas, Aitor Gonzalez-Agirre, Francesco Alessandro Massucci and Marta Villegas ...................... 286

Fair Enough: Standardizing Evaluation and Model Selection for Fairness Research in NLP
Xudong Han, Timothy Baldwin and Trevor Cohn ................................ 297

CHARD: Clinical Health-Aware Reasoning Across Dimensions for Text Generation Models
Steven Y. Feng, Vivek Khetan, Bogdan Sacaleanu, Anatole Gershman and Eduard Hovy .... 313

Prompt Tuning with Contradictory Intentions for Sarcasm Recognition
Yiyi Liu, Ruqing Zhang, Yixing Fan, Jiafeng Guo and Xueqi Cheng ......................... 328

COMBO: A Complete Benchmark for Open KG Canonicalization
Chengyue Jiang, Yong Jiang, Weiqi Wu, Yuting Zheng, Pengjun Xie and Kewei Tu .......... 340

UScore: An Effective Approach to Fully Unsupervised Evaluation Metrics for Machine Translation
Jonas Belouadi and Steffen Eger .............................................. 358

Assistive Recipe Editing through Critiquing
Diego Antognini, Shuyang Li, Boi Faltings and Julian McAuley ............................ 375

DiTTO: A Feature Representation Imitation Approach for Improving Cross-Lingual Transfer
Shanu Kumar, Soujanya Abaraju, Sandipan Dandapat, Sunayana Sitaram and Monojit Choudhury ................................................................. 385

John is 50 years old, can his son be 65? Evaluating NLP Models’ Understanding of Feasibility
Himanshu Gupta, Neeraj Varshney, Swaroop Mishra, Kuntal Kumar Pal, Saurabh Arjun Sawant, Kevin Scaria, Siddharth Goyal and Chitta Baral ........................................ 407

Efficient Encoders for Streaming Sequence Tagging
Ayush Kaushal, Aditya Gupta, Shyam Upadhyay and Manaal Faruqui ................. 418

Retrieve-and-Fill for Scenario-based Task-Oriented Semantic Parsing

Minghao Wu, George Foster, Lizhen Qu and Gholamreza Haffari ....................... 448

Scaling Back-Translation with Domain Text Generation for Sign Language Gloss Translation
Jinhui Ye, Wenxiang Jiao, Xing Wang and Zhaopeng Tu .............................. 463

Realistic Conversational Question Answering with Answer Selection based on Calibrated Confidence and Uncertainty Measurement
Soyeong Jeong, Jinheon Baek, Sung Ju Hwang and Jong Park .......................... 477

xxix
PANCETTA: Phoneme Aware Neural Completion to Elicit Tongue Twisters Automatically
Sedrick Scott Keh, Steven Y. Feng, Varun Gangal, Malihe Alikhani and Eduard Hovy ........... 491

A User-Centered, Interactive, Human-in-the-Loop Topic Modelling System
Zheng Fang, Lama Alqazlan, Du Liu, Yulan He and Rob Procter ......................... 505

A Survey of Methods for Addressing Class Imbalance in Deep-Learning Based Natural Language Processing
Sophie Henning, William Beluch, Alexander Fraser and Annemarie Friedrich .............. 523

Extracting or Guessing? Improving Faithfulness of Event Temporal Relation Extraction
Haoyu Wang, Hongming Zhang, Yuqian Deng, Jacob Gardner, Dan Roth and Muhao Chen .. 541

LoFT: Enhancing Faithfulness and Diversity for Table-to-Text Generation via Logic Form Control
Yilun Zhao, Zhenting Qi, Linyong Nan, Lorenzo Jaime Flores and Dragomir Radev ........ 554

PromptDA: Label-guided Data Augmentation for Prompt-based Few Shot Learners
Canyu Chen and Kai Shu ..................................................... 562

Incorporating Question Answering-Based Signals into Abstractive Summarization via Salient Span Selection
Daniel Deutsch and Dan Roth ............................................. 575

Patient Outcome and Zero-shot Diagnosis Prediction with Hypernetwork-guided Multitask Learning
Shaoxiong Ji and Pekka Marttinen ........................................ 589

A Kind Introduction to Lexical and Grammatical Aspect, with a Survey of Computational Approaches
Annemarie Friedrich, Nianwen Xue and Alexis Palmer ........................................ 599

Incorporating Context into Subword Vocabularies
Shaked Yehezkel and Yuval Pinter ........................................ 623

LoRaLay: A Multilingual and Multimodal Dataset for Long Range and Layout-Aware Summarization
Laura Nguyen, Thomas Scialom, Benjamin Piwowarski and Jacopo Staiano ............... 636

ViHOS: Hate Speech Spans Detection for Vietnamese
Phu Gia Hoang, Canh Duc Luu, Khanh Quoc Tran, Kiet Van Nguyen and Ngan Luu-Thuy Nguyen 652

Vote’n’Rank: Revision of Benchmarking with Social Choice Theory
Mark Rofin, Vladislav Mikhailov, Mikhail Florinsky, Andrey Kravchenko, Tatiana Shavrina, Elena Tutubalina, Daniel Karabekyan and Ekaterina Artemova ................. 670

Combining Parameter-efficient Modules for Task-level Generalisation
Edoardo Maria Ponti, Alessandro Sordoni, Yoshua Bengio and Siva Reddy ................. 687

Self-imitation Learning for Action Generation in Text-based Games
Zijing Shi, Yunqiu Xu, Meng Fang and Ling Chen .............................................. 703

Investigating the Effect of Relative Positional Embeddings on AMR-to-Text Generation with Structural Adapters
Sebastien Montella, Alexis Nasr, Johannes Heinecke, Frederic Bechet and Lina M. Rojas Barahona ................................................................. 727

On the Intersection of Context-Free and Regular Languages
Clemente Pasti, Andreas Opedal, Tiago Pimentel, Tim Vieira, Jason Eisner and Ryan Cotterell 737
Kushal Chawla, Weiyan Shi, Jingwen Zhang, Gale Lucas, Zhou Yu and Jonathan Gratch .... 750

Aggregating Crowdsourced and Automatic Judgments to Scale Up a Corpus of Anaphoric Reference for Fiction and Wikipedia Texts
Juntao Yu, Silviu Paun, Maris Camilleri, Paloma Garcia, Jon Chamberlain, Udo Kruschwitz and Massimo Poesio ................................................................. 767

What Makes Sentences Semantically Related? A Textual Relatedness Dataset and Empirical Study
Mohamed Abdalla, Krishnapriya Vishnubhotla and Saif Mohammad ............ 782

RevUp: Revise and Update Information Bottleneck for Event Representation
Mehdi Rezaee and Francis Ferraro ................................................................. 797

NusaX: Multilingual Parallel Sentiment Dataset for 10 Indonesian Local Languages
Genta Indra Winata, Alham Fikri Aji, Samuel Cakyawijaya, Rahmad Mahendra, Fajri Koto, Ade Romadhony, Kemal Kurniawan, David Moeljadi, Radityo Eko Prasojo, Pascale Fung, Timothy Baldwin, Jey Han Lau, Rico Sennrich and Sebastian Ruder .................................................. 815

The Functional Relevance of Probed Information: A Case Study
Michael Hanna, Roberto Zamparelli and David Mareček .................................... 835

Do Pretrained Contextual Language Models Distinguish between Hebrew Homograph Analyses?
Avi Shmidman, Cheyn Shmuel Shmidman, Dan Bareket, Moshe Koppel and Reut Tsarfaty . 849

Parameter-Efficient Tuning with Special Token Adaptation
Xiaocong Yang, James Y. Huang, Wenhuan Zhou and Muhao Chen .................... 865

Probing Power by Prompting: Harnessing Pre-trained Language Models for Power Connotation Framing
Shima Khanehzar, Trevor Cohn, Gosia Mikolajczak and Lea Frermann .................... 873

Zero and Few-Shot Localization of Task-Oriented Dialogue Agents with a Distilled Representation
Mehrad Moradshahi, Sina Semnani and Monica Lam ........................................ 886

Contextual Semantic Parsing for Multilingual Task-Oriented Dialogues
Mehrad Moradshahi, Victoria Tsai, Giovanni Campagna and Monica Lam ............ 902

Teacher Intervention: Improving Convergence of Quantization Aware Training for Ultra-Low Precision Transformers
Minsoo Kim, Kyuhong Shim, Seongmin Park, Wonyong Sung and Jungwook Choi ........ 916

Generative Replay Inspired by Hippocampal Memory Indexing for Continual Language Learning
Aru Maekawa, Hidetaka Kamigaito, Kotaro Funakoshi and Manabu Okumura ............ 930

A Survey of Multi-task Learning in Natural Language Processing: Regarding Task Relatedness and Training Methods
Zhihan Zhang, Wenhao Yu, Mengxia Yu, Zhichun Guo and Meng Jiang .................... 943

Conclusion-based Counter-Argument Generation
Milad Alshomary and Henning Wachsmuth ..................................................... 957

Question-Answer Sentence Graph for Joint Modeling Answer Selection
Roshni Iyer, Thuy Vu, Alessandro Moschitti and Yizhou Sun ................................ 968

Evaluating and Improving the Coreference Capabilities of Machine Translation Models
Asaf Yehudai, Arie Cattan, Omri Abend and Gabriel Stanovsky ....................... 980

xxxi
Document-Level Planning for Text Simplification
Liam Cripwell, Joël Legrand and Claire Gardent ........................................ 993

Efficient Hybrid Generation Framework for Aspect-Based Sentiment Analysis
Haoran Lv, Junyi Liu, Henan Wang, Yaoming Wang, Jixiang Luo and Yaxiao Liu ...... 1007

What's New? Summarizing Contributions in Scientific Literature
Hiroaki Hayashi, Wojciech Kryscinski, Bryan McCann, Nazneen Rajani and Caiming Xiong 1019

Find Parent then Label Children: A Two-stage Taxonomy Completion Method with Pre-trained Language Model
Fei Xia, Yixuan Weng, Shizhu He, Kang Liu and Jun Zhao .................................. 1032

Meta Self-Refinement for Robust Learning with Weak Supervision
Dawei Zhu, Xiaoyu Shen, Michael Hedderich and Dietrich Klakow .......................... 1043

Looking for a Needle in a Haystack: A Comprehensive Study of Hallucinations in Neural Machine Translation
Nuno M. Guerreiro, Elena Voita and André Martins ........................................... 1059

Investigating UD Treebanks via Dataset Difficulty Measures
Artur Kulmizev and Joakim Nivre ........................................................................... 1076

On Robustness of Prompt-based Semantic Parsing with Large Pre-trained Language Model: An Empirical Study on Codex
Terry Yue Zhuo, Zhuang Li, Yujin Huang, Fatemeh Shiri, Weiqing Wang, Gholamreza Haffari and Yuan-Fang Li ................................................................. 1090

Leveraging Task Dependency and Contrastive Learning for Case Outcome Classification on European Court of Human Rights Cases
Santosh T.Y.S.S, Marcel Perez San Blas, Phillip Kemper and Matthias Grabmair .......... 1103

Semi-supervised Relation Extraction via Data Augmentation and Consistency-training
Komal Teru .............................................................................................................. 1112

Event Temporal Relation Extraction with Bayesian Translational Model
Xingwei Tan, Gabriele Pergola and Yulan He ......................................................... 1125

Persona Expansion with Commonsense Knowledge for Diverse and Consistent Response Generation
Donghyun Kim, Youbin Ahn, Wongyu Kim, Chanhee Lee, Kyungchan Lee, Kyong-Ho Lee, Jeong-guk Kim, Donghoon Shin and Yeonsoo Lee .................................................. 1139

UnifEE: Unified Evidence Extraction for Fact Verification
Nan Hu, Zirui Wu, Yuxuan Lai, Chen Zhang and Yansong Feng .............................. 1150

MiniALBERT: Model Distillation via Parameter-Efficient Recursive Transformers
Mohammadmahdi Nouriborji, Omid Rohanian, Samaneh Kouchaki and David A. Clifton ... 1161

Multilingual Normalization of Temporal Expressions with Masked Language Models
Lukas Lange, Jannik Strötgen, Heike Adel and Dietrich Klakow .................................. 1174

K-hop neighbourhood regularization for few-shot learning on graphs: A case study of text classification
Niels van der Heijden, Ekaterina Shutova and Helen Yannakoudakis .......................... 1187

What Clued the AI Doctor In? On the Influence of Data Source and Quality for Transformer-Based Medical Self-Disclosure Detection
Mina Valizadeh, Xing Qian, Pardis Ranjbar-Noey, Cornelia Caragea and Natalie Parde ... 1201

xxxii
Improving Visual-Semantic Embedding with Adaptive Pooling and Optimization Objective
Zijian Zhang, Chang Shu, Ya Xiao, Yuan Shen, Di Zhu, Youxin Chen, Jing Xiao, Jey Han Lau, Qian Zhang and Zheng Lu ................................................................. 1217

Policy-based Reinforcement Learning for Generalisation in Interactive Text-based Environments
Edan Toledo, Jan Buys and Jonathan Shock ........................................ 1230

Logic Against Bias: Textual Entailment Mitigates Stereotypical Sentence Reasoning
Hongyin Luo and James Glass .............................................................. 1243

Entity Tracking via Effective Use of Multi-Task Learning Model and Mention-guided Decoding
Janvijay Singh, Fan Bai and Zhen Wang ............................................... 1255

Conversational Tree Search: A New Hybrid Dialog Task
Dirk Väth, Lindsey Vanderlyn and Ngoc Thang Vu ................................ 1264

A Human Subject Study of Named Entity Recognition in Conversational Music Recommendation Queries
Elena Epure and Romain Hennequin .................................................... 1281

Entity Disambiguation with Entity Definitions
Luigi Procopio, Simone Conia, Edoardo Barba and Roberto Navigli .......... 1297

Exploring ParaCrawl for Document-level Neural Machine Translation
Yusser Al Ghussin, Jingyi Zhang and Josef van Genabith .......................... 1304

Poor Man’s Quality Estimation: Predicting Reference-Based MT Metrics Without the Reference
Vilém Zouhar, Shehzaad Dhuliawala, Wangchunshu Zhou, Nico Daheim, Tom Kocmi, Yuchen Eleanor Jiang and Mrinmaya Sachan ........................................ 1311

Integrating Translation Memories into Non-Autoregressive Machine Translation
Jitao Xu, Josep Crego and François Yvon ............................................. 1326

Shorten the Long Tail for Rare Entity and Event Extraction
Pengfei Yu and Heng Ji ........................................................................ 1339

Do Deep Neural Networks Capture Compositionality in Arithmetic Reasoning?
Keito Kudo, Yoichi Aoki, Tatsuki Kuribayashi, Ana Brassard, Masashi Yoshikawa, Keisuke Sakaguchi and Kentaro Inui .................................................. 1351

BLM-AgrF: A New French Benchmark to Investigate Generalization of Agreement in Neural Networks
Aixiu An, Chunyang Jiang, Maria A. Rodriguez, Vivi Nastase and Paola Merlo 1363

Robustification of Multilingual Language Models to Real-world Noise in Crosslingual Zero-shot Settings with Robust Contrastive Pretraining
Asa Cooper Stickland, Sailik Sengupta, Jason Krone, Saab Mansour and He He ............................................................. 1375

Unsupervised Anomaly Detection in Multi-Topic Short-Text Corpora
Mira Ait-Saada and Mohamed Nadif ........................................................ 1392

Metaphor Detection with Effective Context Denoising
Shun Wang, Yucheng Li, Chenghua Lin, Loïc Barrault and Frank Guerin .... 1404

Low-Resource Compositional Semantic Parsing with Concept Pretraining
Subendhu Rongali, Mukund Sridhar, Haidar Khan, Konstantine Arkoudas, Wael Hamza and Andrew McCallum ............................................. 1410
Made of Steel? Learning Plausible Materials for Components in the Vehicle Repair Domain
Annerose Eichel, Helena Schlipf and Sabine Schulte im Walde .............................. 1420

Self-Adapted Utterance Selection for Suicidal Ideation Detection in Lifeline Conversations
Zhong-Ling Wang, Po-Hsien Huang, Wen-Yau Hsu and Hen-Hsen Huang ...................... 1436

Can Pretrained Language Models (Yet) Reason Deductively?
Zhangdie Yuan, Songbo Hu, Ivan Vulić, Anna Korhonen and Zaiqiao Meng ..................... 1447

Selective In-Context Data Augmentation for Intent Detection using Pointwise V-Information
Yen-Ting Lin, Alexandros Papangelis, Seokhwan Kim, Sungjin Lee, Devamanyu Hazarika, Mahdi Namazifar, Di Jin, Yang Liu and Dilek Hakkani-Tur .......................... 1463

Multilingual Representation Distillation with Contrastive Learning
Weiting Tan, Kevin Heffernan, Holger Schwenk and Philipp Koehn ......................... 1477

On the inconsistency of separable losses for structured prediction
Caio Corro ........................................ 1491

A Systematic Search for Compound Semantics in Pretrained BERT Architectures
Filip Miletic and Sabine Schulte im Walde .............................................................. 1499

Efficiently Upgrading Multilingual Machine Translation Models to Support More Languages
Simeng Sun, Maha Elbayad, Anna Sun and James Cross ........................................ 1513

Summarize and Generate to Back-translate: Unsupervised Translation of Programming Languages
Wasi Uddin Ahmad, Saikat Chakraborty, Baishakhi Ray and Kai-Wei Chang .................... 1528

The Impacts of Unanswerable Questions on the Robustness of Machine Reading Comprehension Models
Son Quoc Tran, Phong Nguyen-Thuan Do, Uyen Le and Matt Kretchmar ...................... 1543

FrameBERT: Conceptual Metaphor Detection with Frame Embedding Learning
Yucheng Li, Shun Wang, Chenghua Lin, Frank Guerin and Loic Barrault ...................... 1558

Towards More Efficient Insertion Transformer with Fractional Positional Encoding
Zhisong Zhang, Yizhe Zhang and Bill Dolan ............................................................. 1564

SODAPOP: Open-Ended Discovery of Social Biases in Social Commonsense Reasoning Models
Haozhe An, Zongxia Li, Jieryu Zhao and Rachel Rudinger ........................................ 1573

Augmenting Pre-trained Language Models with QA-Memory for Open-Domain Question Answering
Wenhu Chen, Pat Verga, Michiel de Jong, John Wieting and William W. Cohen ............ 1597

Gold Doesn’t Always Glitter: Spectral Removal of Linear and Nonlinear Guarded Attribute Information
Shun Shao, Yftah Ziser and Shay B. Cohen ............................................................... 1611

CTC Alignments Improve Autoregressive Translation
Brian Yan, Siddharth Dalnia, Yosuke Higuchi, Graham Neubig, Florian Metze, Alan W Black and Shinji Watanabe ................................................................. 1623

Modelling Temporal Document Sequences for Clinical ICD Coding
Boon Liang Clarence Ng, Diogo Santos and Marek Rei ............................................. 1640

LongEval: Guidelines for Human Evaluation of Faithfulness in Long-form Summarization
Kalpesh Krishna, Erin Bransom, Bailey Kuehl, Mohit Iyyer, Pradeep Dasigi, Arman Cohan and Kyle Lo .............................................................. 1650
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory-efficient Temporal Moment Localization in Long Videos</td>
<td>1909</td>
</tr>
<tr>
<td>Cristian Rodriguez-Opazo, Edison Marrese-Taylor, Basura Fernando, Hiroya Takamura and Qi Wu</td>
<td></td>
</tr>
<tr>
<td>Extracting Victim Counts from Text</td>
<td>1925</td>
</tr>
<tr>
<td>Mian Zhong, Shehzaad Dhuliawala and Niklas Stoehr</td>
<td></td>
</tr>
<tr>
<td>ConEntail: An Entailment-based Framework for Universal Zero and Few Shot Classification with Supervised Contrastive Pretraining</td>
<td>1941</td>
</tr>
<tr>
<td>Ranran Haoran Zhang, Aysa Xuemo Fan and Rui Zhang</td>
<td></td>
</tr>
<tr>
<td>Guide the Learner: Controlling Product of Experts Debiasing Method Based on Token Attribution Similarities</td>
<td>1954</td>
</tr>
<tr>
<td>Ali Modarressi, Hossein Amirkhani and Mohammad Taher Pilehvar</td>
<td></td>
</tr>
<tr>
<td>Task and Sentiment Adaptation for Appraisal Tagging</td>
<td>1960</td>
</tr>
<tr>
<td>Lin Tian, Xiuzhen Zhang, Myung Hee Kim and Jennifer Biggs</td>
<td></td>
</tr>
<tr>
<td>DREEAM: Guiding Attention with Evidence for Improving Document-Level Relation Extraction</td>
<td>1971</td>
</tr>
<tr>
<td>Youmi Ma, An Wang and Naoaki Okazaki</td>
<td></td>
</tr>
<tr>
<td>Span-based Named Entity Recognition by Generating and Compressing Information</td>
<td>1984</td>
</tr>
<tr>
<td>Nhung T. H. Nguyen, Makoto Miwa and Sophia Ananiadou</td>
<td></td>
</tr>
<tr>
<td>An In-depth Analysis of Implicit and Subtle Hate Speech Messages</td>
<td>1997</td>
</tr>
<tr>
<td>Nicolas Ocampo, Ekaterina Sviridova, Elena Cabrio and Serena Villata</td>
<td></td>
</tr>
<tr>
<td>MTEB: Massive Text Embedding Benchmark</td>
<td>2014</td>
</tr>
<tr>
<td>Niklas Muennighoff, Nouamane Tazi, Loic Magne and Nils Reimers</td>
<td></td>
</tr>
<tr>
<td>Step by Step Loss Goes Very Far: Multi-Step Quantization for Adversarial Text Attacks</td>
<td>2038</td>
</tr>
<tr>
<td>Piotr Gański and Klaudia Balazy</td>
<td></td>
</tr>
<tr>
<td>TwiRGCN: Temporally Weighted Graph Convolution for Question Answering over Temporal Knowledge Graphs</td>
<td>2049</td>
</tr>
<tr>
<td>Aditya Sharma, Apoorv Saxena, Chitrank Gupta, Mehran Kazemi, Partha Talukdar and Soumen Chakrabarti</td>
<td></td>
</tr>
<tr>
<td>ZELDA: A Comprehensive Benchmark for Supervised Entity Disambiguation</td>
<td>2061</td>
</tr>
<tr>
<td>Marcel Milich and Alan Akbik</td>
<td></td>
</tr>
<tr>
<td>GLADIS: A General and Large Acronym Disambiguation Benchmark</td>
<td>2073</td>
</tr>
<tr>
<td>Lihu Chen, Gael Varoquaux and Fabian M. Suchanek</td>
<td></td>
</tr>
<tr>
<td>Probing Cross-Lingual Lexical Knowledge from Multilingual Sentence Encoders</td>
<td>2089</td>
</tr>
<tr>
<td>Ivan Vulić, Goran Glavaš, Fangyu Liu, Nigel Collier, Edoardo Maria Ponti and Anna Korhonen</td>
<td></td>
</tr>
<tr>
<td>Pento-DIARef: A Diagnostic Dataset for Learning the Incremental Algorithm for Referring Expression Generation from Examples</td>
<td>2106</td>
</tr>
<tr>
<td>Philipp Sadler and David Schlangen</td>
<td></td>
</tr>
<tr>
<td>Mitigating Exposure Bias in Grammatical Error Correction with Data Augmentation and Reweighting</td>
<td>2123</td>
</tr>
<tr>
<td>Hannan Cao, Wenmian Yang and Hwee Tou Ng</td>
<td></td>
</tr>
<tr>
<td>Plausible May Not Be Faithful: Probing Object Hallucination in Vision-Language Pre-training</td>
<td>2136</td>
</tr>
<tr>
<td>Wenliang Dai, Zihan Liu, Ziwei Ji, Dan Su and Pascale Fung</td>
<td></td>
</tr>
</tbody>
</table>
Characterizing the Entities in Harmful Memes: Who is the Hero, the Villain, the Victim?
Shivam Sharma, Atharva Kulkarni, Tharun Suresh, Himanshi Mathur, Preslav Nakov, Md. Shad Akhtar and Tanmoy Chakraborty .......................................................... 2149

Systematic Investigation of Strategies Tailored for Low-Resource Settings for Low-Resource Dependency Parsing
Jivnesh Sandhan, Laxmidhar Behera and Pawan Goyal .......................................................... 2164

Compositional Generalisation with Structured Reordering and Fertility Layers
Matthias Lindemann, Alexander Koller and Ivan Titov .......................................................... 2172

Investigating Multi-source Active Learning for Natural Language Inference
Ard Snijders, Douwe Kiela and Katerina Margatina .......................................................... 2187

Towards a Unified Multi-Domain Multilingual Named Entity Recognition Model
Mayank Kulkarni, Daniel Preotiuc-Pietro, Karthik Radhakrishnan, Genta Indra Winata, Shijie Wu, Lingjie Xie and Shaohua Yang .......................................................... 2210

Do Neural Topic Models Really Need Dropout? Analysis of the Effect of Dropout in Topic Modeling
Suman Adhya, Avishek Lahiri and Debarshi Kumar Sanyal .......................................................... 2220

A Psycholinguistic Analysis of BERT’s Representations of Compounds
Lars Buijtenlaar and Sandro Pezzelle .......................................................... 2230

Measuring Normative and Descriptive Biases in Language Models Using Census Data
Samia Touileb, Lilja Øvrelid and Erik Velldal .......................................................... 2242

UDAPTER - Efficient Domain Adapation Using Adapters
Bhavitvya Malik, Abhina Ramesh Kashyap, Min-Yen Kan and Soujanya Poria .......................................................... 2249

Efficient CTC Regularization via Coarse Labels for End-to-End Speech Translation
Biao Zhang, Barry Haddow and Rico Sennrich .......................................................... 2264

Exploring Category Structure with Contextual Language Models and Lexical Semantic Networks
Joseph Renner, Pascal Denis, Remi Gilleron and Angèle Brunelliêre .......................................................... 2277

An Empirical Study of Clinical Note Generation from Doctor-Patient Encounters
Asma Ben Abacha, Wen-wai Yim, Yadan Fan and Thomas Lin .......................................................... 2291

Instruction Clarification Requests in Multimodal Collaborative Dialogue Games: Tasks, and an Analysis of the CoDraw Dataset
Brielen Madureira and David Schlangen .......................................................... 2303

Can Synthetic Text Help Clinical Named Entity Recognition? A Study of Electronic Health Records in French
Nicolas Hiebel, Olivier Ferret, Karen Fort and Aurélie Névéol .......................................................... 2320

IRMA: the 335-million-word Italian coRpus for studying MisinformAtion
Fabio Carrella, Alessandro Miani and Stephan Lewandowsky .......................................................... 2339

Parameter-Efficient Korean Character-Level Language Modeling
Marco Cognetta, Sangwhan Moon, Lawrence Wolf-sonkin and Naoaki Okazaki .......................................................... 2350

Opportunities and Challenges in Neural Dialog Tutoring
Jakub Macina, Nico Daheim, Lingzhi Wang, Tanmay Sinha, Manu Kapur, Iryna Gurevych and Mrinmaya Sachan .......................................................... 2357
Penguins Don’t Fly: Reasoning about Generics through Instantiations and Exceptions
Emily Allaway, Jena D. Hwang, Chandra Bhagavatula, Kathleen McKeown, Doug Downey and Yejin Choi ................................................................. 2618

Adding Instructions during Pretraining: Effective way of Controlling Toxicity in Language Models
Shrimai Prabhumoye, Mostofa Patwary, Mohammad Shoeybi and Bryan Catanzaro ............ 2636

Multi2Claim: Generating Scientific Claims from Multi-Choice Questions for Scientific Fact-Checking
Neset Tan, Trung Nguyen, Josh Bensemann, Alex Peng, Qiming Bao, Yang Chen, Mark Gahegan and Michael Witbrock .............................................................................. 2652

On Evaluation of Document Classifiers using RVL-CDIP
Stefan Larson, Gordon Lim and Kevin Leach ................................................................. 2665

Event Linking: Grounding Event Mentions to Wikipedia
Xiaodong Yu, Wenpeng Yin, Nitish Gupta and Dan Roth .............................................. 2679

SwitchPrompt: Learning Domain-Specific Gated Soft Prompts for Classification in Low-Resource Domains
Koustava Goswami, Lukas Lange, Jun Araki and Heike Adel ........................................ 2689

Do dialogue representations align with perception? An empirical study
Sarenne Wallbridge, Peter Bell and Catherine Lai ....................................................... 2696

Methods for Measuring, Updating, and Visualizing Factual Beliefs in Language Models
Peter Hase, Mona Diab, Asli Celikyilmaz, Xian Li, Zornitsa Kozareva, Veselin Stoyanov, Mohit Bansal and Srinivasan Iyer .................................................. 2714

Improving Sign Recognition with Phonology
Lee Kezar, Jesse Thomason and Zed Sehyr ................................................................. 2732

Parameter-efficient Modularised Bias Mitigation via AdapterFusion
Deepak Kumar, Oleg Lesota, George Zerveas, Daniel Cohen, Carsten Eickhoff, Markus Schedl and Navid Rekabsaz ...................................................... 2738

LingMess: Linguistically Informed Multi Expert Scorers for Coreference Resolution
Shon Otmazgin, Arie Cattan and Yoav Goldberg ...................................................................... 2752

Finding the Law: Enhancing Statutory Article Retrieval via Graph Neural Networks
Antoine Louis, Gijs van Dijck and Gerasimos Spanakis .............................................. 2761

Behavior Cloned Transformers are Neurosymbolic Reasoners
Ruoyao Wang, Peter Jansen, Marc-Alexandre Côté and Prithviraj Ammanabrolu ............ 2777

Bridging the Gap Between BabelNet and HowNet: Unsupervised Sense Alignment and Sememe Prediction
Xiang Zhang, Ning Shi, Bradley Hauer and Grzegorz Kondrak ...................................... 2789

The StatCan Dialogue Dataset: Retrieving Data Tables through Conversations with Genuine Intents
Xing Han Lu, Siva Reddy and Harm de Vries ............................................................... 2799

Question Generation Using Sequence-to-Sequence Model with Semantic Role Labels
Ali reza Naejji, Aijun An, Heidar Davoudi, Marjan Delpisheh and Muath Alzghool .......... 2830

StyLEx: Explaining Style Using Human Lexical Annotations
Shirley Anugrah Hayati, Kyumin Park, Dheeraj Rajagopal, Lyle Ungar and Dongyeop Kang 2843
Comparing Intrinsic Gender Bias Evaluation Measures without using Human Annotated Examples
Masahiro Kaneko, Danushka Bollegala and Naoaki Okazaki ........................................2857

Faithfulness-Aware Decoding Strategies for Abstractive Summarization
David Wan, Mengwen Liu, Kathleen McKeown, Markus Dreyer and Mohit Bansal ........2864

Dynamic Benchmarking of Masked Language Models on Temporal Concept Drift with Multiple Views
Katerina Margatina, Shuai Wang, Yogarshi Vyas, Neha Anna John, Yassine Benajiba and Miguel Ballesteros ........................................2881

Anjana Arunkumar, Swaroop Mishra, Bhavdeep Singh Sachdeva, Chitta Baral and Chris Bryan 2899

COMPS: Conceptual Minimal Pair Sentences for testing Robust Property Knowledge and its Inheritance in Pre-trained Language Models
Kanishka Misra, Julia Rayz and Allyson Ettinger .........................................................2928

Probabilistic Robustness for Data Filtering
Yu Yu, Abdul Rafae Khan, Shahram Khadivi and Jia Xu ...........................................2950

Unsupervised Improvement of Factual Knowledge in Language Models
Nafis Sadeq, Byungkyu Kang, Prarit Lamba and Julian McAuley ..............................2960

Learning to Ignore Adversarial Attacks
Yiming Zhang, Yangqiaoyu Zhou, Samuel Carton and Chenhao Tan .........................2970

Should You Mask 15% in Masked Language Modeling?
Alexander Wettig, Tianyu Gao, Zexuan Zhong and Danqi Chen .................................2985

How do Words Contribute to Sentence Semantics? Revisiting Sentence Embeddings with a Perturbation Method
Wenlin Yao, Lifeng Jin, Hongming Zhang, Xiaoman Pan, Kaiqiang Song, Dian Yu, Dong Yu and Jianshu Chen .................................................3001

AutoTriggER: Label-Efficient and Robust Named Entity Recognition with Auxiliary Trigger Extraction
Dong-Ho Lee, Ravi Kiran Selvam, Sheikh Muhammad Sarwar, Bill Yuchen Lin, Fred Morstatter, Jay Pujara, Elizabeth Boschee, James Allan and Xiang Ren ...........................................3011

Incorporating Task-Specific Concept Knowledge into Script Learning
Chenkai Sun, Tie Xu, ChengXiang Zhai and Heng Ji ...................................................3026

DeepMaven: Deep Question Answering on Long-Distance Movie/TV Show Videos with Multimedia Knowledge Extraction and Synthesis
Yi Fung, Han Wang, Tong Wang, Ali Kebarighotbi, Mohit Bansal, Heng Ji and Prem Natarajan 3041

Salient Span Masking for Temporal Understanding
Jeremy R. Cole, Aditi Chaudhary, Bhuwan Dhingra and Partha Talukdar ......................3052

PECO: Examining Single Sentence Label Leakage in Natural Language Inference Datasets through Progressive Evaluation of Cluster Outliers
Michael Saxon, Xinyi Wang, Wenda Xu and William Yang Wang ..............................3061

Weakly-Supervised Questions for Zero-Shot Relation Extraction
Saeed Najafi and Alona Fyshe ......................................................................................3075
DiffQG: Generating Questions to Summarize Factual Changes
Jeremy R. Cole, Palak Jain, Julian Martin Eisenschlos, Michael J.Q. Zhang, Eunsol Choi and Bhuwan Dhingra ................................................................. 3088

Contextual Dynamic Prompting for Response Generation in Task-oriented Dialog Systems
Sandesh Swamy, Narges Tabari, Chacha Chen and Rashmi Gangadharaiah ................................. 3102

Why Can’t Discourse Parsing Generalize? A Thorough Investigation of the Impact of Data Diversity
Yang Janet Liu and Amir Zeldes ................................................................. 3112

Enriching Biomedical Knowledge for Low-resource Language Through Large-scale Translation
Long Phan, Tai Dang, Hieu Tran, Trieu H. Trinh, Vy Phan, Lam D. Chau and Minh-Thang Luong 3113

Syntax-guided Neural Module Distillation to Probe Compositionality in Sentence Embeddings
Rohan Pandey .................................................. ........................................ 3143

Closed-book Question Generation via Contrastive Learning
Xiangjue Dong, Jiaying Lu, Jianling Wang and James Caverlee .................................................. 3150

A Hybrid Detection and Generation Framework with Separate Encoders for Event Extraction
Ge Shi, Yunyue Su, Yongliang Ma and Ming Zhou .................................................. ........................................ 3163

Path Spuriousness-aware Reinforcement Learning for Multi-Hop Knowledge Graph Reasoning
Chunyang Jiang, Tianchen Zhu, Haoyi Zhou, Chang Liu, Ting Deng, Chunming Hu and Jianxin Li ................................................................. 3181

Self-Adaptive Named Entity Recognition by Retrieving Unstructured Knowledge
Kosuke Nishida, Naoki Yoshinaga and Kyosuke Nishida ................................................................. 3193

When Do Pre-Training Biases Propagate to Downstream Tasks? A Case Study in Text Summarization
Faisal Ladhak, Esin Durmus, Mirac Suzgun, Tianyi Zhang, Dan Jurafsky, Kathleen McKeown and Tatsunori Hashimoto .................................................. ........................................ 3206

BERT Shows Garden Path Effects
Tovah Irwin, Kyra Wilson and Alec Marantz ................................................................. 3220

Models Teaching Models: Improving Model Accuracy with Slingshot Learning
Lachlan O’Neill, Nandini Anantharama, Satya Borgohain and Simon D. Angus ........................................ 3233

A Federated Approach for Hate Speech Detection
Jay Gala, Deep Gandhi, Jash Mehta and Zeerak Talat ................................................................. 3248

Learning the Legibility of Visual Text Perturbations
Dev Seth, Rickard Stureborg, Danish Pruthi and Bhuwan Dhingra .................................................. 3260

DyLoRA: Parameter-Efficient Tuning of Pre-trained Models using Dynamic Search-Free Low-Rank Adaptation
Mojtaba Valipour, Mehdi Rezagholizadeh, Ivan Kobyzev and Ali Ghodsi .................................................. 3274

Conversational Emotion-Cause Pair Extraction with Guided Mixture of Experts
DongJin Jeong and JinYeong Bak ................................................................. 3288

Language Generation Models Can Cause Harm: So What Can We Do About It? An Actionable Survey
Sachin Kumar, Vidhisha Balachandran, Lucille Njoo, Antonios Anastasopoulos and Yulia Tsvetkov ................................................................. 3299
Keng Ji Chow, Samson Tan and Min-Yen Kan .................................................. 3322

Paraphrase Acquisition from Image Captions
Marcel Gohsen, Matthias Hagen, Martin Potthast and Benno Stein .................. 3348

Generation-Based Data Augmentation for Offensive Language Detection: Is It Worth It?
Camilla Casula and Sara Tonelli ................................................................. 3359

Quantifying Context Mixing in Transformers
Hosein Mohebbi, Willem Zuidema, Grzegorz Chrupała and Afra Alishahi ........... 3378

KGVL-BART: Knowledge Graph Augmented Visual Language BART for Radiology Report Generation
Kaveri Kale, Pushpak Bhattacharyya, Milind Gune, Aditya Shetty and Rustom Lawyer .... 3401

A simple but effective model for attachment in discourse parsing with multi-task learning for relation labeling
Zineb Bennis, Julie Hunter and Nicholas Asher ............................................... 3412

How Far Can It Go? On Intrinsic Gender Bias Mitigation for Text Classification
Ewoenam Kwaku Tokpo, Pieter Delobelle, Bettina Berendt and Toon Calders .......... 3418

Multimodal Event Transformer for Image-guided Story Ending Generation
Yucheng Zhou and Guodong Long ............................................................... 3434

Improving Cross-modal Alignment for Text-Guided Image Inpainting
Yucheng Zhou and Guodong Long ............................................................... 3445

Semantic Specialization for Knowledge-based Word Sense Disambiguation
Sakae Mizuki and Naoaki Okazaki .............................................................. 3457

Concept-based Persona Expansion for Improving Diversity of Persona-Grounded Dialogue
Donghyun Kim, Youbin Ahn, Chanhee Lee, Wongyu Kim, Kyong-Ho Lee, Donghoon Shin and Yeonsoo Lee ................................................................. 3471

RPTCS: A Reinforced Persona-aware Topic-guiding Conversational System
Zishan Ahmad, Kshitij Mishra, Asif Ekbal and Pushpak Bhattacharyya .................. 3482

What Did You Learn To Hate? A Topic-Oriented Analysis of Generalization in Hate Speech Detection
Tom Bourgeade, Patricia Chiril, Farah Benamara and Véronique Moriceau ........... 3495

End-to-end Case-Based Reasoning for Commonsense Knowledge Base Completion
Zonglin Yang, Xinya Du, Erik Cambria and Claire Cardie ............................... 3509

Marwa Gaser, Manuel Mager, Injy Hamed, Nizar Habash, Slim Abdennadher and Ngoc Thang Vu ................................................................. 3523

Identifying the limits of transformers when performing model-checking with natural language
Tharindu Madusanka, Riza Batista-navarro and Ian Pratt-hartmann .................... 3539

Improving the Generalizability of Collaborative Dialogue Analysis With Multi-Feature Embeddings
Ayesha Enayet and Gita Sukthankar .......................................................... 3551

MetaQA: Combining Expert Agents for Multi-Skill Question Answering
Haritz Puerto, Gözde Şahin and Iryna Gurevych .......................................... 3566
DiscoScore: Evaluating Text Generation with BERT and Discourse Coherence
Wei Zhao, Michael Strube and Steffen Eger ......................................................... 3865

Know your audience: specializing grounded language models with listener subtraction
Aaditya K Singh, David Ding, Andrew Saxe, Felix Hill and Andrew Lampinen .............. 3884

Meeting the Needs of Low-Resource Languages: The Value of Automatic Alignments via Pretrained Models