

18th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2024)

Volume 1 (Long Papers)

St. Julian's, Malta
17 - 22 March 2024

Part 1 of 4

ISBN: 978-1-7138-9442-1

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2024) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Enhancing Ethical Explanations of Large Language Models through Iterative Symbolic Refinement</i> Xin Quan, Marco Valentino, Louise A. Dennis and Andre Freitas	1
<i>Multi-Relational Hyperbolic Word Embeddings from Natural Language Definitions</i> Marco Valentino, Danilo Carvalho and Andre Freitas	23
<i>Anisotropy Is Inherent to Self-Attention in Transformers</i> Nathan Godey, Éric Villemonte De La Clergerie and Benoît Sagot	35
<i>Generating Benchmarks for Factuality Evaluation of Language Models</i> Dor Muhlgay, Ori Ram, Inbal Magar, Yoav Levine, Nir Ratner, Yonatan Belinkov, Omri Abend, Kevin Leyton-Brown, Amnon Shashua and Yoav Shoham	49
<i>Leak, Cheat, Repeat: Data Contamination and Evaluation Malpractices in Closed-Source LLMs</i> Simone Balloccu, Patrícia Schmidtová, Mateusz Lango and Ondrej Dusek	67
<i>Archer: A Human-Labeled Text-to-SQL Dataset with Arithmetic, Commonsense and Hypothetical Reasoning</i> Danna Zheng, Mirella Lapata and Jeff Z. Pan	94
<i>GEAR: Augmenting Language Models with Generalizable and Efficient Tool Resolution</i> Yining Lu, Haoping Yu and Daniel Khashabi	112
<i>LLM Comparative Assessment: Zero-shot NLG Evaluation through Pairwise Comparisons using Large Language Models</i> Adian Liusie, Potsawee Manakul and Mark Gales	139
<i>Parameter-Efficient Conversational Recommender System as a Language Processing Task</i> Mathieu Ravaut, Hao Zhang, Lu Xu, Aixin Sun and Yong Liu	152
<i>OpenPI2.0: An Improved Dataset for Entity Tracking in Texts</i> Li Zhang, Hainiu Xu, Abhinav Kommula, Chris Callison-Burch and Niket Tandon	166
<i>A Comparative Multidimensional Analysis of Empathetic Systems</i> Andrew Lee, Jonathan K. Kummerfeld, Larry Ann and Rada Mihalcea	179
<i>Few-Shot Data Synthesis for Open Domain Multi-Hop Question Answering</i> Mingda Chen, Xilun Chen and Wen-tau Yih	190
<i>Language Models as Inductive Reasoners</i> Zonglin Yang, Li Dong, Xinya Du, Hao Cheng, Erik Cambria, Xiaodong Liu, Jianfeng Gao and Furu Wei	209
<i>SIB-200: A Simple, Inclusive, and Big Evaluation Dataset for Topic Classification in 200+ Languages and Dialects</i> David Ifeoluwa Adelani, Hannah Liu, Xiaoyu Shen, Nikita Vassilyev, Jesujoba Oluwadara Alabi, Yanke Mao, Haonan Gao and En-Shiun Annie Lee	226
<i>FinBPM: A Framework for Portfolio Management-based Financial Investor Behavior Perception Model</i> Zhilu Zhang, Procheta Sen, Zimu Wang, Ruoyu Sun, Zhengyong Jiang and Jionglong Su	246
<i>Asking the Right Question at the Right Time: Human and Model Uncertainty Guidance to Ask Clarification Questions</i> Alberto Testoni and Raquel Fernández	258

<i>Like a Good Nearest Neighbor: Practical Content Moderation and Text Classification</i> Luke Bates and Iryna Gurevych	276
<i>Zero-shot Sentiment Analysis in Low-Resource Languages Using a Multilingual Sentiment Lexicon</i> Fajri Koto, Tilman Beck, Zeerak Talat, Iryna Gurevych and Timothy Baldwin	298
<i>CEAN: Contrastive Event Aggregation Network with LLM-based Augmentation for Event Extraction</i> Zihao Meng, Tao Liu, Heng Zhang, Kai Feng and Peng Zhao	321
<i>How Transferable are Attribute Controllers on Pretrained Multilingual Translation Models?</i> Danni Liu and Jan Niehues	334
<i>MultiMUC: Multilingual Template Filling on MUC-4</i> William Gantt, Shabnam Behzad, Hannah YoungEun An, Yunmo Chen, Aaron Steven White, Benjamin Van Durme and Mahsa Yarmohammadi	349
<i>Align and Augment: Generative Data Augmentation for Compositional Generalization</i> Francesco Cazzaro, Davide Locatelli and Ariadna Quattoni	369
<i>UNSEE: Unsupervised Non-contrastive Sentence Embeddings</i> Ömer Veysel Çağatan	384
<i>EXPLORER: Exploration-guided Reasoning for Textual Reinforcement Learning</i> Kinjal Basu, Keerthiram Murugesan, Subhajt Chaudhury, Murray Campbell, Kartik Talamadupula and Tim Klinger	394
<i>From Text Segmentation to Smart Chaptering: A Novel Benchmark for Structuring Video Transcriptions</i> Fabian Retkowski and Alexander Waibel	406
<i>Fréchet Distance for Offline Evaluation of Information Retrieval Systems with Sparse Labels</i> Negar Arabzadeh and Charles L. A. Clarke	420
<i>Semantic Sensitivities and Inconsistent Predictions: Measuring the Fragility of NLI Models</i> Erik Arakelyan, Zhaoqi Liu and Isabelle Augenstein	432
<i>Exploring the Robustness of Task-oriented Dialogue Systems for Colloquial German Varieties</i> Ekaterina Artemova, Verena Blaschke and Barbara Plank	445
<i>PEARL: Prompting Large Language Models to Plan and Execute Actions Over Long Documents</i> Simeng Sun, Yang Liu, Shuohang Wang, Dan Iter, Chenguang Zhu and Mohit Iyyer	469
<i>LAraBench: Benchmarking Arabic AI with Large Language Models</i> Ahmed Abdelali, Hamdy Mubarak, Shammur Absar Chowdhury, Maram Hasanain, Basel Mousi, Sabri Boughorbel, Samir Abdaljalil, Yassine El Kheir, Daniel Izham, Fahim Dalvi, Majd Hawasly, Nizi Nazar, Youssef Ibrahim Elshahawy, Ahmed Ali, Nadir Durrani, Natasa Milic-Frayling and Firoj Alam 487	
<i>SentenceLDA: Discriminative and Robust Document Representation with Sentence Level Topic Model</i> Taehun Cha and Donghun Lee	521
<i>Towards Hierarchical Spoken Language Disfluency Modeling</i> Jiachen Lian and Gopala Anumanchipalli	539
<i>Finding a Needle in the Adversarial Haystack: A Targeted Paraphrasing Approach For Uncovering Edge Cases with Minimal Distribution Distortion</i> Aly M. Kassem and Sherif Saad	552

<i>FAIR: Filtering of Automatically Induced Rules</i> Divya Jyoti Bajpai, Ayush Maheshwari, Manjesh Kumar Hanawal and Ganesh Ramakrishnan	573
<i>NNOSE: Nearest Neighbor Occupational Skill Extraction</i> Mike Zhang, Rob Van Der Goot, Min-Yen Kan and Barbara Plank	589
<i>GAINER: Graph Machine Learning with Node-specific Radius for Classification of Short Texts and Documents</i> Naganand Yadati	609
<i>MAFIA: Multi-Adapter Fused Inclusive Language Models</i> Prachi Jain, Ashutosh Sathe, Varun Gumma, Kabir Ahuja and Sunayana Sitaram	627
<i>Code-Switched Language Identification is Harder Than You Think</i> Laurie Burchell, Alexandra Birch, Robert Peter Thompson and Kenneth Heafield	646
<i>Generation-driven Contrastive Self-training for Zero-shot Text Classification with Instruction-following LLM</i> Ruohong Zhang, Yau-Shian Wang and Yiming Yang	659
<i>Quantifying the Hyperparameter Sensitivity of Neural Networks for Character-level Sequence-to-Sequence Tasks</i> Adam Wiemerslage, Kyle Gorman and Katharina Von Der Wense	674
<i>Examining Gender and Racial Bias in Large Vision–Language Models Using a Novel Dataset of Parallel Images</i> Kathleen C. Fraser and Svetlana Kiritchenko	690
<i>ConstraintChecker: A Plugin for Large Language Models to Reason on Commonsense Knowledge Bases</i> Quyut V. Do, Tianqing Fang, Shizhe Diao, Zhaowei Wang and Yangqiu Song	714
<i>A* shortest string decoding for non-idempotent semirings</i> Kyle Gorman and Cyril Allauzen	732
<i>Importance-Aware Data Augmentation for Document-Level Neural Machine Translation</i> Minghao Wu, Yufei Wang, George Foster, Lizhen Qu and Gholamreza Haffari	740
<i>Lost in Translationese? Reducing Translation Effect Using Abstract Meaning Representation</i> Shira Wein and Nathan Schneider	753
<i>Comparing Template-based and Template-free Language Model Probing</i> Sagi Shaier, Kevin Bennett, Lawrence Hunter and Katharina Von Der Wense	766
<i>Desiderata For The Context Use Of Question Answering Systems</i> Sagi Shaier, Lawrence Hunter and Katharina Von Der Wense	777
<i>Scaling up Discovery of Latent Concepts in Deep NLP Models</i> Majd Hawasly, Fahim Dalvi and Nadir Durrani	793
<i>AnthroScore: A Computational Linguistic Measure of Anthropomorphism</i> Myra Cheng, Kristina Gligoric, Tiziano Piccardi and Dan Jurafsky	807
<i>Centering the Speech Community</i> Steven Bird and Dean Yibarbuk	826
<i>Improving the TENOR of Labeling: Re-evaluating Topic Models for Content Analysis</i> Zongxia Li, Andrew Mao, Daniel Kofi Stephens, Pranav Goel, Emily Walpole, Alden Dima, Juan Francisco Fung and Jordan Lee Boyd-Graber	840

<i>Quality Does Matter: A Detailed Look at the Quality and Utility of Web-Mined Parallel Corpora</i> Surangika Ranathunga, Nisansa De Silva, Velayuthan Menan, Aloka Fernando and Charitha S.m. Rathnayake	860
<i>VOLTAGE: A Versatile Contrastive Learning based OCR Methodology for ultra low-resource scripts through Auto Glyph Feature Extraction</i> Prawaal Sharma, Poonam Goyal, Vidisha Sharma and Navneet Goyal	881
<i>Unsupervised Contrast-Consistent Ranking with Language Models</i> Niklas Stoehr, Pengxiang Cheng, Jing Wang, Daniel Preotiuc-Pietro and Rajarshi Bhowmik .	900
<i>Entity-level Factual Adaptiveness of Fine-tuning based Abstractive Summarization Models</i> Jongyoon Song, Nohil Park, Bongkyu Hwang, Jaewoong Yun, Seongho Joe, Youngjune Gwon and Sungroh Yoon	915
<i>Meme-ingful Analysis: Enhanced Understanding of Cyberbullying in Memes Through Multimodal Explanations</i> Prince Jha, Krishanu Maity, Raghav Jain, Apoorv Verma, Sriparna Saha and Pushpak Bhattacharyya	930
<i>LaMini-LM: A Diverse Herd of Distilled Models from Large-Scale Instructions</i> Minghao Wu, Abdul Waheed, Chiyu Zhang, Muhammad Abdul-Mageed and Alham Fikri Aji	944
<i>Automated Cognate Detection as a Supervised Link Prediction Task with Cognate Transformer</i> V.S.D.S.Mahesh Akavarapu and Arnab Bhattacharya	965
<i>Leveraging Multi-lingual Positive Instances in Contrastive Learning to Improve Sentence Embedding</i> Kaiyan Zhao, Qiyu Wu, Xin-Qiang Cai and Yoshimasa Tsuruoka	976
<i>Moderation in the Wild: Investigating User-Driven Moderation in Online Discussions</i> Neele Falk, Eva Maria Vecchi, Iman Jundi and Gabriella Lapesa	992
<i>Cross-Lingual Transfer from Related Languages: Treating Low-Resource Maltese as Multilingual Code-Switching</i> Kurt Micallef, Nizar Habash, Claudia Borg, Fadhl Eryani and Houda Bouamor	1014
<i>Where Do We Go From Here? Multi-scale Allocentric Relational Inference from Natural Spatial Descriptions</i> Tzuf Paz-Argaman, John Palowitch, Sayali Kulkarni, Jason Michael Baldrige and Reut Tsarfaty	1026
<i>Bias in Opinion Summarisation from Pre-training to Adaptation: A Case Study in Political Bias</i> Nannan Huang, Haytham M. Fayek and Xiuzhen Zhang	1041
<i>Document Structure in Long Document Transformers</i> Jan Buchmann, Max Eichler, Jan-Micha Bodensohn, Ilia Kuznetsov and Iryna Gurevych ...	1056
<i>The Role of Data Curation in Image Captioning</i> Wenyan Li, Jonas F. Lotz, Chen Qiu and Desmond Elliott	1074
<i>Large-Scale Bitext Corpora Provide New Evidence for Cognitive Representations of Spatial Terms</i> Peter Viechnicki, Kevin Duh, Anthony Kostacos and Barbara Landau	1089
<i>REFINER: Reasoning Feedback on Intermediate Representations</i> Debjit Paul, Mete Ismayilzada, Maxime Peyrard, Beatriz Borges, Antoine Bosselut, Robert West and Boi Faltings	1100

<i>HumbEL: A Human-in-the-Loop Approach for Evaluating Demographic Factors of Language Models in Human-Machine Conversations</i>	
Anthony Sicilia, Jennifer C. Gates and Malihe Alikhani	1127
<i>LOCOST: State-Space Models for Long Document Abstractive Summarization</i>	
Florian Le Bronnec, Song Duong, Mathieu Ravaut, Alexandre Allauzen, Nancy F. Chen, Vincent Guigue, Alberto Lumbreras, Laure Soulier and Patrick Gallinari	1144
<i>A Classification-Guided Approach for Adversarial Attacks against Neural Machine Translation</i>	
Sahar Sadrizadeh, Ljiljana Dolamic and Pascal Frossard	1160
<i>Improving Generalization in Semantic Parsing by Increasing Natural Language Variation</i>	
Irina Saparina and Mirella Lapata	1178
<i>Text-to-Code Generation with Modality-relative Pre-training</i>	
Fenia Christopoulou, Guchun Zhang and Gerasimos Lampouras	1194
<i>No Error Left Behind: Multilingual Grammatical Error Correction with Pre-trained Translation Models</i>	
Agnes Luhtaru, Elizaveta Korotkova and Mark Fishel	1209
<i>Quantifying Stereotypes in Language</i>	
Yang Liu	1223
<i>Generation, Distillation and Evaluation of Motivational Interviewing-Style Reflections with a Foundational Language Model</i>	
Andrew Brown, Jiading Zhu, Mohamed Abdelwahab, Alec Dong, Cindy Wang and Jonathan Rose	1241
<i>Multi-Reference Benchmarks for Russian Grammatical Error Correction</i>	
Frank Palma Gomez and Alla Rozovskaya	1253
<i>Plan-Grounded Large Language Models for Dual Goal Conversational Settings</i>	
Diogo Glória-Silva, Rafael Ferreira, Diogo Tavares, David Semedo and Joao Magalhaes . . .	1271
<i>“Define Your Terms” : Enhancing Efficient Offensive Speech Classification with Definition</i>	
Huy Nghiem, Umang Gupta and Fred Morstatter	1293
<i>VlogQA: Task, Dataset, and Baseline Models for Vietnamese Spoken-Based Machine Reading Comprehension</i>	
Thinh Phuoc Ngo, Khoa Tran Anh Dang, Son T. Luu, Kiet Van Nguyen and Ngan Luu-Thuy Nguyen	1310
<i>CEV-LM: Controlled Edit Vector Language Model for Shaping Natural Language Generations</i>	
Samraj Moorjani, Adit Krishnan and Hari Sundaram	1325
<i>It’s All Relative: Learning Interpretable Models for Scoring Subjective Bias in Documents from Pairwise Comparisons</i>	
Aswin Suresh, Wu Chi Hsuan and Matthias Grossglauser	1341
<i>HiGen: Hierarchy-Aware Sequence Generation for Hierarchical Text Classification</i>	
Vidit Jain, Mukund Rungta, Yuchen Zhuang, Yue Yu, Zeyu Wang, Mu Gao, Jeffrey Skolnick and Chao Zhang	1354
<i>M4: Multi-generator, Multi-domain, and Multi-lingual Black-Box Machine-Generated Text Detection</i>	
Yuxia Wang, Jonibek Mansurov, Petar Ivanov, Jinyan Su, Artem Shelmanov, Akim Tsvigun, Chenxi Whitehouse, Osama Mohammed Afzal, Tarek Mahmoud, Toru Sasaki, Thomas Arnold, Alham Fikri Aji, Nizar Habash, Iryna Gurevych and Preslav Nakov	1369

<i>A Truly Joint Neural Architecture for Segmentation and Parsing</i> Danit Yshaayahu Levi and Reut Tsarfaty	1408
<i>ViLexNorm: A Lexical Normalization Corpus for Vietnamese Social Media Text</i> Thanh-Nhi Nguyen, Thanh-Phong Le and Kiet Van Nguyen	1421
<i>Diffusion-NAT: Self-Prompting Discrete Diffusion for Non-Autoregressive Text Generation</i> Kun Zhou, Yifan Li, Xin Zhao and Ji-Rong Wen	1438
<i>Unleashing the Power of Discourse-Enhanced Transformers for Propaganda Detection</i> Alexander Chernyavskiy, Dmitry Ilvovsky and Preslav Nakov	1452
<i>Predicting Client Emotions and Therapist Interventions in Psychotherapy Dialogues</i> Tobias Mayer, Neha Warikoo, Amir Eliassaf, Dana Atzil-Slonim and Iryna Gurevych	1463
<i>Who Needs Decoders? Efficient Estimation of Sequence-Level Attributes with Proxies</i> Yassir Fathullah, Puria Radmard, Adian Liusie and Mark Gales	1478
<i>3D Rotation and Translation for Hyperbolic Knowledge Graph Embedding</i> Yihua Zhu and Hidetoshi Shimodaira	1497
<i>Geo-Encoder: A Chunk-Argument Bi-Encoder Framework for Chinese Geographic Re-Ranking</i> Yong Cao, Ruixue Ding, Boli Chen, Xianzhi Li, Min Chen, Daniel Hershcovich, Pengjun Xie and Fei Huang	1516
<i>Style-News: Incorporating Stylized News Generation and Adversarial Verification for Neural Fake News Detection</i> Wei-Yao Wang, Yu-Chieh Chang and Wen-Chih Peng	1531
<i>Graph-based Clustering for Detecting Semantic Change Across Time and Languages</i> Xianghe Ma, Michael Strube and Wei Zhao	1542
<i>Translate to Disambiguate: Zero-shot Multilingual Word Sense Disambiguation with Pretrained Language Models</i> Haoqiang Kang, Terra Blevins and Luke Zettlemoyer	1562
<i>Anchor Points: Benchmarking Models with Much Fewer Examples</i> Rajan Pathe Vivek, Kawin Ethayarajh, Diyi Yang and Douwe Kiela	1576
<i>SCO-VIST: Social Interaction Commonsense Knowledge-based Visual Storytelling</i> Eileen Wang, Caren Han and Josiah Poon	1602
<i>Discovering and Articulating Frames of Communication from Social Media Using Chain-of-Thought Reasoning</i> Maxwell Weinzierl and Sanda Harabagiu	1617
<i>VEIL: Vetting Extracted Image Labels from In-the-Wild Captions for Weakly-Supervised Object Detection</i> Arushi Rai and Adriana Kovashka	1632
<i>WSC+: Enhancing The Winograd Schema Challenge Using Tree-of-Experts</i> Pardis Sadat Zahraei and Ali Emami	1650
<i>Kardeş-NLU: Transfer to Low-Resource Languages with Big Brother’s Help – A Benchmark and Evaluation for Turkic Languages</i> Lütfi Kerem Senel, Benedikt Ebing, Konul Baghirova, Hinrich Schuetze and Goran Glavaš	1672

<i>Chaining Event Spans for Temporal Relation Grounding</i>	
Jongho Kim, Dohyeon Lee, Minsoo Kim and Seung-won Hwang	1689
<i>Fine-Grained Natural Language Inference Based Faithfulness Evaluation for Diverse Summarisation Tasks</i>	
Huajian Zhang, Yumo Xu and Laura Perez-Beltrachini	1701
<i>AnaDE1.0: A Novel Data Set for Benchmarking Analogy Detection and Extraction</i>	
Bhavya Bhavya, Shradha Sehgal, Jinjun Xiong and ChengXiang Zhai	1723
<i>A Comprehensive Survey of Sentence Representations: From the BERT Epoch to the CHATGPT Era and Beyond</i>	
Abhinav Ramesh Kashyap, Thanh-Tung Nguyen, Viktor Schlegel, Stefan Winkler, See-Kiong Ng and Soujanya Poria	1738
<i>Learning to Retrieve In-Context Examples for Large Language Models</i>	
Liang Wang, Nan Yang and Furu Wei	1752
<i>EnCore: Fine-Grained Entity Typing by Pre-Training Entity Encoders on Coreference Chains</i>	
Frank Martin Mtumbuka and Steven Schockaert	1768
<i>Unsupervised stance detection for social media discussions: A generic baseline</i>	
Maia Sutter, Antoine Gourru, Amine Trabelsi and Christine Largeron	1782
<i>Putting Context in Context: the Impact of Discussion Structure on Text Classification</i>	
Nicolò Penzo, Antonio Longa, Bruno Lepri, Sara Tonelli and Marco Guerini	1793
<i>Aligning Large and Small Language Models via Chain-of-Thought Reasoning</i>	
Leonardo Ranaldi and Andre Freitas	1812
<i>Disentangling the Roles of Target-side Transfer and Regularization in Multilingual Machine Translation</i>	
Yan Meng and Christof Monz	1828
<i>Uncovering Stereotypes in Large Language Models: A Task Complexity-based Approach</i>	
Hari Shrawgi, Prasanjit Rath, Tushar Singhal and Sandipan Dandapat	1841
<i>Rainbow - A Benchmark for Systematic Testing of How Sensitive Visio-Linguistic Models are to Color Naming</i>	
Marie Bexte, Andrea Horbach and Torsten Zesch	1858
<i>CATfOOD: Counterfactual Augmented Training for Improving Out-of-Domain Performance and Calibration</i>	
Rachneet Singh Sachdeva, Martin Tutek and Iryna Gurevych	1876
<i>UP5: Unbiased Foundation Model for Fairness-aware Recommendation</i>	
Wenyue Hua, Yingqiang Ge, Shuyuan Xu, Jianchao Ji, Zelong Li and Yongfeng Zhang	1899
<i>Human Temporal Inferences Go Beyond Aspectual Class</i>	
Katarzyna Pruś, Mark Steedman and Adam Lopez	1913
<i>It is not True that Transformers are Inductive Learners: Probing NLI Models with External Negation</i>	
Michael Sullivan	1924
<i>Polarized Opinion Detection Improves the Detection of Toxic Language</i>	
John Pavlopoulos and Aristidis Likas	1946

<i>Improving Acoustic Word Embeddings through Correspondence Training of Self-supervised Speech Representations</i>	Amit Meghanani and Thomas Hain	1959
<i>Investigating Agency of LLMs in Human-AI Collaboration Tasks</i>	Ashish Sharma, Sudha Rao, Chris Brockett, Akanksha Malhotra, Nebojsa Jojic and Bill Dolan	1968
<i>SynthDST: Synthetic Data is All You Need for Few-Shot Dialog State Tracking</i>	Atharva Kulkarni, Bo-Hsiang Tseng, Joel Ruben Antony Moniz, Dhivya Piraviperumal, Hong Yu and Shruti Bhargava	1988
<i>Argument Mining as a Text-to-Text Generation Task</i>	Masayuki Kawarada, Tsutomu Hirao, Wataru Uchida and Masaaki Nagata	2002
<i>Answering legal questions from laymen in German civil law system</i>	Marius Büttner and Ivan Habernal	2015
<i>An Empirical Analysis of Diversity in Argument Summarization</i>	Michiel Van Der Meer, Piek Vossen, Catholijn M Jonker and Pradeep Kumar Murukannaiah	2028
<i>What Makes Medical Claims (Un)Verifiable? Analyzing Entity and Relation Properties for Fact Verification</i>	Amelie Wuehrl, Yarik Menchaca Resendiz, Lara Grimminger and Roman Klinger	2046
<i>Approximate Attributions for Off-the-Shelf Siamese Transformers</i>	Lucas Moeller, Dmitry Nikolaev and Sebastian Padó	2059
<i>Describing Images Fast and Slow: Quantifying and Predicting the Variation in Human Signals during Visuo-Linguistic Processes</i>	Ece Takmaz, Sandro Pezzelle and Raquel Fernández	2072
<i>Tracing the Roots of Facts in Multilingual Language Models: Independent, Shared, and Transferred Knowledge</i>	Xin Zhao, Naoki Yoshinaga and Daisuke Oba	2088
<i>Comparing Knowledge Sources for Open-Domain Scientific Claim Verification</i>	Juraj Vladika and Florian Matthes	2103
<i>Measuring Uncertainty in Neural Machine Translation with Similarity-Sensitive Entropy</i>	Julius Cheng and Andreas Vlachos	2115
<i>LegalLens: Leveraging LLMs for Legal Violation Identification in Unstructured Text</i>	Dor Bernsohn, Gil Semo, Yaron Vazana, Gila Hayat, Ben Hagag, Joel Niklaus, Rohit Saha and Kyril Truskovskiy	2129
<i>μPLAN: Summarizing using a Content Plan as Cross-Lingual Bridge</i>	Fantine Huot, Joshua Maynez, Chris Alberti, Reinald Kim Amplayo, Priyanka Agrawal, Constanza Fierro, Shashi Narayan and Mirella Lapata	2146
<i>Exploring Data Augmentation in Neural DRS-to-Text Generation</i>	Muhammad Saad Amin, Luca Anselma and Alessandro Mazzei	2164
<i>Think Twice: Measuring the Efficiency of Eliminating Prediction Shortcuts of Question Answering Models</i>	Lukáš Mikula, Michal Štefánik, Marek Petrovič and Petr Sojka	2179

<i>Improving Contrastive Learning in Emotion Recognition in Conversation via Data Augmentation and Decoupled Neutral Emotion</i>	
Yujin Kang and Yoon-Sik Cho	2194
<i>Neuralign: A Context-Aware, Cross-Lingual and Fully-Neural Sentence Alignment System for Long Texts</i>	
Francesco Maria Molfese, Andrei Stefan Bejgu, Simone Tedeschi, Simone Conia and Roberto Navigli.....	2209
<i>Explaining Speech Classification Models via Word-Level Audio Segments and Paralinguistic Features</i>	
Eliana Pastor, Alkis Koudounas, Giuseppe Attanasio, Dirk Hovy and Elena Baralis	2221
<i>Zero-Shot End-to-End Spoken Language Understanding via Cross-Modal Selective Self-Training</i>	
Jianfeng He, Julian Salazar, Kaisheng Yao, Haoqi Li and Jason Cai	2239
<i>Clever Hans or Neural Theory of Mind? Stress Testing Social Reasoning in Large Language Models</i>	
Natalie Shapira, Mosh Levy, Seyed Hossein Alavi, Xuhui Zhou, Yejin Choi, Yoav Goldberg, Maarten Sap and Vered Shwartz	2257
<i>NevIR: Negation in Neural Information Retrieval</i>	
Orion Weller, Dawn Lawrie and Benjamin Van Durme	2274
<i>“According to . . . ”: Prompting Language Models Improves Quoting from Pre-Training Data</i>	
Orion Weller, Marc Marone, Nathaniel Weir, Dawn Lawrie, Daniel Khashabi and Benjamin Van Durme	2288
<i>Accurate and Well-Calibrated ICD Code Assignment Through Attention Over Diverse Label Embeddings</i>	
Goncalo Emanuel Cavaco Gomes, Isabel Pereira Coutinho and Bruno Martins	2302
<i>Investigating Content Planning for Navigating Trade-offs in Knowledge-Grounded Dialogue</i>	
Kushal Chawla, Hannah Rashkin, Gaurav Singh Tomar and David Reitter.....	2316
<i>SPUQ: Perturbation-Based Uncertainty Quantification for Large Language Models</i>	
Xiang Gao, Jiaxin Zhang, Lalla Mouatadid and Kamalika Das.....	2336
<i>TESS: Text-to-Text Self-Conditioned Simplex Diffusion</i>	
Rabeeh Karimi Mahabadi, Hamish Ivison, Jaesung Tae, James Henderson, Iz Beltagy, Matthew E Peters and Arman Cohan.....	2347
<i>Advancing Precise Outline-Conditioned Text Generation with Task Duality and Explicit Outline Control</i>	
Yunzhe Li, Qian Chen, Weixiang Yan, Wen Wang, Qinglin Zhang and Hari Sundaram	2362
<i>Localization vs. Semantics: Visual Representations in Unimodal and Multimodal Models</i>	
Zhuowan Li, Cihang Xie, Benjamin Van Durme and Alan Yuille.....	2378
<i>Creating Suspenseful Stories: Iterative Planning with Large Language Models</i>	
Kaige Xie and Mark Riedl	2391
<i>Few-Shot Dialogue Summarization via Skeleton-Assisted Prompt Transfer in Prompt Tuning</i>	
Kaige Xie, Tong Yu, Haoliang Wang, Junda Wu, Handong Zhao, Ruiyi Zhang, Kanak Mahadik, Ani Nenkova and Mark Riedl.....	2408
<i>Ask, Assess, and Refine: Rectifying Factual Consistency and Hallucination in LLMs with Metric-Guided Feedback Learning</i>	
Dongyub Lee, Eunhwan Park, Hodong Lee and Heuseok Lim.....	2422

<i>Effective Controllable Bias Mitigation for Classification and Retrieval using Gate Adapters</i>	
Shahed Masoudian, Cornelia Volaucnik, Markus Schedl and Navid Rekabsaz	2434
<i>STable: Table Generation Framework for Encoder-Decoder Models</i>	
Michał Pietruszka, Michał Turski, Łukasz Borchmann, Tomasz Dwojak, Gabriela Nowakowska, Karolina Szyndler, Dawid Jurkiewicz and Łukasz Garncarek	2454
<i>A RelEntLess Benchmark for Modelling Graded Relations between Named Entities</i>	
Asahi Ushio, Jose Camacho-Collados and Steven Schockaert	2473
<i>A Multimodal Framework to Detect Target Aware Aggression in Memes</i>	
Shawly Ahsan, Eftekhari Hossain, Omar Sharif, Avishek Das, Mohammed Moshul Hoque and M. Ali Akber Dewan	2487
<i>Graph Guided Question Answer Generation for Procedural Question-Answering</i>	
Hai X. Pham, Isma Hadji, Xinnuo Xu, Ziedune Degutyte, Jay Rainey, Evangelos Kazakos, Afsaneh Fazly, Georgios Tzimiropoulos and Brais Martinez	2501
<i>Contrastive Decoding Reduces Hallucinations in Large Multilingual Machine Translation Models</i>	
Jonas Waldendorf, Barry Haddow and Alexandra Birch	2526
<i>Leveraging fine-tuned Large Language Models with LoRA for Effective Claim, Claimer, and Claim Object Detection</i>	
Sotiris Kotitsas, Panagiotis Kounoudis, Eleni Koutli and Haris Papageorgiou	2540
<i>Should I try multiple optimizers when fine-tuning a pre-trained Transformer for NLP tasks? Should I tune their hyperparameters?</i>	
Nefeli Gkouti, Prodromos Malakasiotis, Stavros Toumpis and Ion Androutsopoulos	2555
<i>GUMsley: Evaluating Entity Saliency in Summarization for 12 English Genres</i>	
Jessica Lin and Amir Zeldes	2575
<i>Sensitivity, Performance, Robustness: Deconstructing the Effect of Sociodemographic Prompting</i>	
Tilman Beck, Hendrik Schuff, Anne Lauscher and Iryna Gurevych	2589
<i>Threat Behavior Textual Search by Attention Graph Isomorphism</i>	
Chanwoo Bae, Guan hong Tao, Zhuo Zhang and Xiangyu Zhang	2616
<i>Identifying Narrative Content in Podcast Transcripts</i>	
Yosra Abdessamed, Shadi Rezapour and Steven R. Wilson	2631
<i>Frequency Explains the Inverse Correlation of Large Language Models' Size, Training Data Amount, and Surprisal's Fit to Reading Times</i>	
Byung-Doh Oh, Shisen Yue and William Schuler	2644
<i>Presentations by the Humans and For the Humans: Harnessing LLMs for Generating Persona-Aware Slides from Documents</i>	
Ishani Mondal, Shwetha S, Anandhavelu Natarajan, Aparna Garimella, Sambaran Bandyopadhyay and Jordan Lee Boyd-Graber	2664
<i>ToPro: Token-Level Prompt Decomposition for Cross-Lingual Sequence Labeling Tasks</i>	
Bolei Ma, Ercong Nie, Shuzhou Yuan, Helmut Schmid, Michael Färber, Frauke Kreuter and Hinrich Schuetze	2685
<i>Small Language Models Improve Giants by Rewriting Their Outputs</i>	
Giorgos Vernikos, Arthur Brazinskas, Jakub Adamek, Jonathan Mallinson, Aliaksei Severyn and Eric Malmi	2703

<i>Unintended Bias Detection and Mitigation in Misogynous Memes</i>	
Gitanjali Kumari, Anubhav Sinha and Asif Ekbal	2719
<i>A Weak Supervision Approach for Few-Shot Aspect Based Sentiment Analysis</i>	
Robert Vacareanu, Siddharth Varia, Kishaloy Halder, Shuai Wang, Giovanni Paolini, Neha Anna John, Miguel Ballesteros and Smaranda Muresan	2734
<i>Counterfactual Reasoning with Knowledge Graph Embeddings</i>	
Lena Zellinger, Andreas Stephan and Benjamin Roth	2753
<i>System-Level Natural Language Feedback</i>	
Weizhe Yuan, Kyunghyun Cho and Jason E Weston	2773
<i>Syntactic Preposing and Discourse Relations</i>	
Yunfang Dong, Xixian Liao and Bonnie L. Webber	2790
<i>Can we obtain significant success in RST discourse parsing by using Large Language Models?</i>	
Aru Maekawa, Tsutomu Hirao, Hidetaka Kamigaito and Manabu Okumura	2803
<i>Ameli: Enhancing Multimodal Entity Linking with Fine-Grained Attributes</i>	
Barry Menglong Yao, Sijia Wang, Yu Chen, Qifan Wang, Minqian Liu, Zhiyang Xu, Licheng Yu and Lifu Huang	2816
<i>Generative Dense Retrieval: Memory Can Be a Burden</i>	
Peiwen Yuan, Xinglin Wang, Shaoxiong Feng, Boyuan Pan, Yiwei Li, Heda Wang, Xupeng Miao and Kan Li	2835
<i>Backward Compatibility During Data Updates by Weight Interpolation</i>	
Raphael Schumann, Elman Mansimov, Yi-An Lai, Nikolaos Pappas, Xibin Gao and Yi Zhang	2846
<i>Gradient-Based Language Model Red Teaming</i>	
Nevan Wichers, Carson Denison and Ahmad Beirami	2862
<i>Do Moral Judgment and Reasoning Capability of LLMs Change with Language? A Study using the Multilingual Defining Issues Test</i>	
Aditi Khandelwal, Utkarsh Agarwal, Kumar Tanmay and Monojit Choudhury	2882
<i>Analyzing the Evaluation of Cross-Lingual Knowledge Transfer in Multilingual Language Models</i>	
Sara Rajae and Christof Monz	2895
<i>Large-Scale Label Interpretation Learning for Few-Shot Named Entity Recognition</i>	
Jonas Golde, Felix Hamborg and Alan Akbik	2915
<i>MLCopilot: Unleashing the Power of Large Language Models in Solving Machine Learning Tasks</i>	
Lei Zhang, Yuge Zhang, Kan Ren, Dongsheng Li and Yuqing Yang	2931
<i>Text-Guided Image Clustering</i>	
Andreas Stephan, Lukas Miklautz, Kevin Sidak, Jan Philip Wahle, Bela Gipp, Claudia Plant and Benjamin Roth	2960
<i>CCPrefix: Counterfactual Contrastive Prefix-Tuning for Many-Class Classification</i>	
Yang Li, Canran Xu, Guodong Long, Tao Shen, Chongyang Tao and Jing Jiang	2977