

19th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2026)

Volume 1 (Long Papers)

St. Julian's, Malta
24-29 March 2026

Part 1 of 12

ISBN: 979-8-3313-3449-9

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© (2026) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

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>LM-Lexicon: Improving Definition Modeling via Harmonizing Semantic Experts</i> Yang Liu, Jiaye Yang, Weikang Li, Jiahui Liang, Yang Li and Lingyong Yan	1
<i>Teams of LLM Agents can Exploit Zero-Day Vulnerabilities</i> Yuxuan Zhu, Antony Kellermann, Akul Gupta, Philip Li, Richard Fang, Rohan Bindu and Daniel Kang	23
<i>Can Reasoning Help Large Language Models Capture Human Annotator Disagreement?</i> Jingwei Ni, Yu Fan, Vilém Zouhar, Donya Rooein, Alexander Miserlis Hoyle, Mrinmaya Sachan, Markus Leippold, Dirk Hovy and Elliott Ash	36
<i>Early-Exit and Instant Confidence Translation Quality Estimation</i> Vilém Zouhar, Maike Züfle, Beni Egressy, Julius Cheng, Mrinmaya Sachan and Jan Niehues .	55
<i>Simplifying Outcome of Language Model Component Analyses with ELIA</i> A. Eidt, N. Felhus.	77
<i>SCoPE VLM: Selective Context Processing for Efficient Document Navigation in Vision-Language Models</i> Gyubeum Lim, Yemo Koo and Vijay Krishna Madiseti	95
<i>Beyond Sample-Level Feedback: Using Reference-Level Feedback to Guide Data Synthesis</i> Shuhaib Mehri, Xiusi Chen, Heng Ji and Dilek Hakkani-Tür	141
<i>T2-RAGBench: Text-and-Table Aware Retrieval-Augmented Generation</i> Jan Strich, Enes Kutay Isgorur, Maximilian Trescher, Chris Biemann and Martin Semmann .	165
<i>The Pragmatic Mind of Machines: Tracing the Emergence of Pragmatic Competence in Large Language Models</i> Kefan Yu, Qingcheng Zeng, Weihao Xuan, Wanxin Li, Jingyi Wu and Rob Voigt	192
<i>Hierarchical Text Classification with LLM-Refined Taxonomies</i> Jonas Golde, Nicolaas Paul Jedema, RaviKiran Krishnan and Phong Le	214
<i>Divide, Reweight, and Conquer: A Logit Arithmetic Approach for In-Context Learning</i> Chengsong Huang, Langlin Huang and Jiaxin Huang	229
<i>Understanding Jailbreak Success: A Study of Latent Space Dynamics in Large Language Models</i> Sarah Ball, Frauke Kreuter and Nina Panickssery	250
<i>Out of Style: RAG’s Fragility to Linguistic Variation</i> Tianyu Cao, Neel Bhandari, Akhila Yerukola, Akari Asai and Maarten Sap	280
<i>Do Political Opinions Transfer Between Western Languages? An Analysis of Unaligned and Aligned Multilingual LLMs</i> Franziska Weeber, Tanise Ceron and Sebastian Padó	319
<i>H-MEM: Hierarchical Memory for High-Efficiency Long-Term Reasoning in LLM Agents</i> Haoran Sun, Shaoning Zeng and Bob Zhang	341
<i>MULSUM: A Multimodal Summarization System with Vis-Aligner and Diversity-Aware Image Selection</i> Abid Ali, Diego Molla and Usman Naseem	351

<i>How Quantization Shapes Bias in Large Language Models</i> Federico Marcuzzi, Xuefei Ning, Roy Schwartz and Iryna Gurevych	363
<i>If Probable, Then Acceptable? Understanding Conditional Acceptability Judgments in Large Language Models</i> Jasmin Orth, Philipp Mondorf and Barbara Plank	405
<i>The Dog the Cat Chased Stumped the Model: Measuring When Language Models Abandon Structure for Shortcuts</i> Sangmitra Madhusudan, Kaige Chen and Ali Emami	428
<i>Automated Screening of Antibacterial Nanoparticle Literature: Dataset Curation and Model Evaluation</i> Alperen Ozturk, Şaziye Betül Özateş, Sophia Bahar Root, Angela Violi, Nicholas Kotov, J. Scott VanEpps and Emine Sumeysra Turali Emre	454
<i>Intention Knowledge Graph Construction for User Intention Relation Modeling</i> Jiaxin Bai, Zhaobo Wang, Junfei Cheng, Dan Yu, Zerui Huang, Weiqi Wang, Xin Liu, Chen Luo, Yanming Zhu, Bo Li and Yangqiu Song	466
<i>Analogical Structure, Minimal Contextual Cues and Contrastive Distractors: Input Design for Sample-Efficient Linguistic Rule Induction</i> Chunyang Jiang and Paola Merlo	485
<i>JiraiBench: A Bilingual Benchmark for Evaluating Large Language Models' Detection of Human risky health behavior Content in Jirai Community</i> Yunze Xiao, Tingyu He, Lionel Z. Wang, Yiming Ma, Xingyu Song, Xiaohang Xu, Mona T. Diab, Irene Li and Ka Chung Ng	501
<i>Chandomitra: Towards Generating Structured Sanskrit Poetry from Natural Language Inputs</i> Manoj Balaji Jagadeeshan, Samarth Bhatia, Pretam Ray, Harshul Raj Surana, Akhil Rajeev P, Priya Mishra, Annarao Kulkarni, Ganesh Ramakrishnan, Prathosh AP and Pawan Goyal	518
<i>Tailored Emotional LLM-Supporter: Enhancing Cultural Sensitivity</i> Chen Cecilia Liu, Hiba Arnaout, Nils Kovačić, Dana Atzil-Slonim and Iryna Gurevych	535
<i>Leveraging LLM-GNN Integration for Open-World Question Answering over Knowledge Graphs</i> Hussein Abdallah, Ibrahim Abdelaziz, Panos Kalnis and Essam Mansour	575
<i>Democratic or Authoritarian? Probing a New Dimension of Political Biases in Large Language Models</i> David Guzman Piedrahita, Irene Strauss, Rada Mihalcea and Zhijing Jin	593
<i>PromptFE: Automated Feature Engineering by Prompting</i> Yufeng Zou, Jean Utke, Diego Klabjan and Han Liu	653
<i>Detecting (Un)answerability in Large Language Models with Linear Directions</i> Maor Juliet Lavi, Tova Milo and Mor Geva	682
<i>Online Difficulty Filtering for Reasoning Oriented Reinforcement Learning</i> Sanghwan Bae, Jiwoo Hong, Min Young Lee, Hanbyul Kim, Jeongyeon Nam and Donghyun Kwak	700
<i>BERT, are you paying attention? Attention regularization with human-annotated rationales</i> Elize Herrewijnen, Dong Nguyen, Floris Bex and Albert Gatt	720
<i>Humans and transformer LMs: Abstraction drives language learning</i> Jasper Jian and Christopher D Manning	752

<i>BigTokDetect: A Clinically-Informed Vision–Language Modeling Framework for Detecting Pro-Bigorexia Videos on TikTok</i>	
Minh Duc Chu, Kshitij Pawar, Zihao He, Roxanna Sharifi, Ross M. Sonnenblick, Magdalayna Curry, Laura DAdamo, Lindsay Young, Stuart Murray and Kristina Lerman	766
<i>Do language models accommodate their users? A study of linguistic convergence</i>	
Terra Blevins, Susanne Schmalwieser and Benjamin Roth	791
<i>Auditing Language Model Unlearning via Information Decomposition</i>	
Anmol Goel, Alan Ritter and Iryna Gurevych	808
<i>OD-Stega: LLM-Based Relatively Secure Steganography via Optimized Distributions</i>	
Yu-Shin Huang, Peter Just, Hanyun Yin, Krishna Narayanan, Ruihong Huang and Chao Tian	827
<i>Sparse Adapter Fusion for Continual Learning in NLP</i>	
Min Zeng, Xi Chen, Haiqin Yang and Yike Guo	852
<i>Rethinking Prompt Optimizers: From Prompt Merits to Optimization</i>	
Zixiao Zhu, Hanzhang Zhou, Zijian Feng, Tianjiao Li, Chua Jia Jim Deryl, Lee Onn Mak, Gee Wah Ng and Kezhi Mao	864
<i>A Survey on Multilingual Mental Disorders Detection from Social Media Data</i>	
Ana-Maria Bucur, Marcos Zampieri, Tharindu Ranasinghe and Fabio Crestani	893
<i>Identifying Fine-grained Forms of Populism in Political Discourse: A Case Study on Donald Trump’s Presidential Campaigns</i>	
Ilias Chalkidis, Stephanie Brandl and Paris Aslanidis	919
<i>SCoNE: a Self-Correcting and Noise-Augmented Method for Complex Biological and Chemical Named Entity Recognition</i>	
Xingyu Zhu, Claire Nédellec, Balazs Nagy, Laszlo Vidacs and Robert Bossy	937
<i>A Benchmark for Audio Reasoning Capabilities of Multimodal Large Language Models</i>	
Iwona Christop, Mateusz Czyżnikiewicz, Paweł Skórzewski, Łukasz Bondaruk, Jakub Kubiak, Marcin Lewandowski and Marek Kubis	953
<i>Nemotron-CrossThink: Scaling Self-Learning beyond Math Reasoning</i>	
Syeda Nahida Akter, Shrimai Prabhumoye, Matvei Novikov, Seungju Han, Ying Lin, Evelina Bakhturina, Eric Nyberg, Yejin Choi, Mostofa Patwary, Mohammad Shoeybi and Bryan Catanzaro	984
<i>Safety of Large Language Models Beyond English: A Systematic Literature Review of Risks, Biases, and Safeguards</i>	
Aleksandra Krasnodębska, Katarzyna Dziewulska, Karolina Seweryn, Maciej Chrabaszcz and Wojciech Kusa	1003
<i>InfiGUIAgent: A Multimodal Generalist GUI Agent with Native Reasoning and Reflection</i>	
Yuhang Liu, Pengxiang Li, Zishu Wei, Congkai Xie, Xueyu Hu, Xinchun Xu, Shengyu Zhang, Xiaotian Han, Hongxia Yang and Fei Wu	1035
<i>Cetvel: A Unified Benchmark for Evaluating Language Understanding, Generation and Cultural Capacity of LLMs for Turkish</i>	
Yakup Abrek Er, Ilker Kesen, Gözde Gül Şahin and Aykut Erdem	1052
<i>CALE : Concept-Aligned Embeddings for Both Within-Lemma and Inter-Lemma Sense Differentiation</i>	
Bastien Liétard and Gabriel Loiseau	1086

<i>Do NOT Classify and Count: Hybrid Attribute Control Success Evaluation</i>	
Felix Matthias Saaro, Pius von Däniken, Mark Cieliebak and Jan Milan Deriu	1101
<i>Detecting Training Data of Large Language Models via Expectation Maximization</i>	
Gyuwan Kim, Yang Li, Evangelia Spiliopoulou, Jie Ma and William Yang Wang	1115
<i>How effective are VLMs in assisting humans in inferring the quality of mental models from Multimodal short answers?</i>	
Pritam Sil, Durgaprasad Karnam, Vinay Reddy Venumuddala and Pushpak Bhattacharyya	1130
<i>Don't Trust Generative Agents to Mimic Communication on Social Networks Unless You Benchmarked their Empirical Realism</i>	
Simon Münker, Nils Schwager and Achim Rettinger	1141
<i>Persona Prompting as a Lens on LLM Social Reasoning</i>	
Jing Yang, Moritz Hechtbauer, Elisabeth Khalilov, Evelyn Luise Brinkmann, Vera Schmitt and Nils Feldhus	1152
<i>PartisanLens: A Multilingual Dataset of Hyperpartisan and Conspiratorial Immigration Narratives in European Media</i>	
Michele Joshua Maggini, Paloma Piot, Anxo Pérez, Erik Bran Marino, Lúa Santamaría Montesinos, Ana Lisboa Cotovio, Marta Vázquez Abuín, Javier Parapar and Pablo Gamallo	1171
<i>Adaptive LLM-Symbolic Reasoning via Dynamic Logical Solver Composition</i>	
Lei Xu, Pierre Beckmann, Marco Valentino and Andre Freitas	1187
<i>Lexical Popularity: Quantifying the Impact of Pre-training for LLM Performance</i>	
Elena Sofia Ruzzetti, Fabio Massimo Zanzotto and Tommaso Caselli	1209
<i>Uncovering Hidden Correctness in LLM Causal Reasoning via Symbolic Verification</i>	
Paul He, Yinya Huang, Mrinmaya Sachan and Zhijing Jin	1231
<i>CORE: Measuring Multi-Agent LLM Interaction Quality under Game-Theoretic Pressures</i>	
Punya Syon Pandey, Yongjin Yang, Jiarui Liu and Zhijing Jin	1251
<i>Aleph-Alpha-GermanWeb: Improving German-language LLM pre-training with model-based data curation and synthetic data generation</i>	
Thomas F Burns, Letitia Parcalabescu, Stephan Waeldchen, Michael Barlow, Gregor Zieglertrum, Volker Stampa, Bastian Harren and Björn Deiseroth	1267
<i>Ultra-Low-Dimensional Prompt Tuning via Random Projection</i>	
Zijun Wu, Yongchang Hao and Lili Mou	1284
<i>NP-Hard Lower Bound Complexity for Semantic Self-Verification</i>	
Robin Young	1304
<i>STAMP: Selective Task-Aware Mechanism for Text Privacy</i>	
Fengwei Tian, Payel Bhattacharjee, Heidi Hanson, Geoffrey D Rubin, Joseph Y. Lo and Ravi Tandon	1319
<i>Deconstructing Instruction-Following: A New Benchmark for Granular Evaluation of Large Language Model Instruction Compliance Abilities</i>	
Alberto Purpura, Li Wang, Sahil Badyal, Gene Beaufrand and Adam Faulkner	1334
<i>Utterance-level Detection Framework for LLM-Involved Content Detection in Conversational Setting</i>	
Muyang Zhou and Huaxia Rui	1350

<i>Patient-Similarity Cohort Reasoning in Clinical Text-to-SQL</i> Yifei Shen, Yilun Zhao, Justice Ou, Tinglin Huang and Arman Cohan	1367
<i>iBERT: Interpretable Embeddings via Sense Decomposition</i> Vishal Anand, Milad Alshomary and Kathleen McKeown	1413
<i>Attacker’s Noise Can Manipulate Your Audio-based LLM in the Real World</i> Vinu Sankar Sadasivan, Soheil Feizi, Rajiv Mathews and Lun Wang	1430
<i>Say It Another Way: Auditing LLMs with a User-Grounded Automated Paraphrasing Framework</i> Clea Chataigner, Rebecca Ma, Prakhar Ganesh, Yuhao Chen, Afaf Taik, Elliot Creager and Golnoosh Farnadi	1441
<i>AutoBool: Reinforcement-Learned LLM for Effective Automatic Systematic Reviews Boolean Query Generation</i> Shuai Wang, Harrisen Scells, Bevan Koopman and Guido Zuccon	1468
<i>Improving LLM Domain Certification with Pretrained Guide Models</i> Jiaqian Zhang, Zhaozhi Qian, Faroq AL-Tam, Ignacio Iacobacci, Muhammad AL-Qurishi and Riad Souissi	1494
<i>TDFlow: Agentic Workflows for Test Driven Development</i> Kevin Han, Siddharth Maddikayala, Tim Knappe, Om Patel, Austen Liao and Amir Barati Farmani	1511
<i>Contrastive Learning with Narrative Twins for Modeling Story Salience</i> Igor Sterner, Alex Lascarides and Frank Keller	1528
<i>ExAnte: A Benchmark for Ex-Ante Inference in Large Language Models</i> Yachuan Liu, Xiaochun Wei, Lin Shi, Xinnuo Li, Bohan Zhang, Paramveer Dhillon and Qiaozhu Mei	1551
<i>CRADLE Bench: A Clinician-Annotated Benchmark for Multi-Faceted Mental Health Crisis and Safety Risk Detection</i> Grace Byun, Rebecca Lipschutz, Sean T. Minton, Abigail Powers and Jinho D. Choi	1572
<i>Coordinates from Context: Using LLMs to Ground Complex Location References</i> Tessa Masis and Brendan O’Connor	1591
<i>Discourse Graph Guided Document Translation with Large Language Models</i> Viet Thanh Pham, Minghan Wang, Hao-Han Liao and Thuy-Trang Vu	1607
<i>StarFlow: Generating Structured Workflow Outputs From Sketch Images</i> Patrice Bechard, Chao Wang, Amirhossein Abaskohi, Juan A. Rodriguez, Christopher Pal, David Vazquez, Spandana Gella, Sai Rajeswar and Perouz Taslakian	1628
<i>Adaptive Helpfulness–Harmlessness Alignment with Preference Vectors</i> Ren-Wei Liang, Chin Ting Hsu, Chan-Hung Yu, Saransh Agrawal, Shih-Cheng Huang, Chieh-Yen Lin, Shang-Tse Chen, Kuan-Hao Huang and Shao-Hua Sun	1646
<i>How Reliable are Confidence Estimators for Large Reasoning Models? A Systematic Benchmark on High-Stakes Domains</i> Reza Khanmohammadi, Erfan Miahi, Simerjot Kaur, Charese Smiley, Ivan Brugere, Kundan S Thind and Mohammad M. Ghassemi	1669
<i>SearchLLM: Detecting LLM Paraphrased Text by Measuring the Similarity with Regeneration of the Candidate Source via Search Engine</i> Hoang-Quoc Nguyen-Son, Minh-Son Dao and Koji Zettsu	1755

<i>RoZO: Geometry-Aware Zeroth-Order Fine-Tuning on Low-Rank Adapters for Black-Box Large Language Models</i>	
Zichen Song and Weijia Li	1773
<i>Mitigating Degree Bias in Hypergraphs via Attribute-as-Structure Approach</i>	
Ryusei Nishide and Makoto Miwa	1784
<i>Generative Personality Simulation via Theory-Informed Structured Interview</i>	
Pengda Wang, Huiqi Zou, Han Jiang, Hanjie Chen, Tianjun Sun, Xiaoyuan Yi, Ziang Xiao and Frederick L. Oswald	1802
<i>Unraveling LLM Jailbreaks Through Safety Knowledge Neurons</i>	
Chongwen Zhao, Yutong Ke and Kaizhu Huang	1889
<i>ELLA: Efficient Lifelong Learning for Adapters in Large Language Models</i>	
Shristi Das Biswas, Yue Zhang, Anwesan Pal, Radhika Bhargava and Kaushik Roy	1907
<i>LingGen: Scalable Multi-Attribute Linguistic Control via Power-Law Masking</i>	
Mohamed Elgaar and Hadi Amiri	1925
<i>RECIPE-TKG: From Sparse History to Structured Reasoning for LLM-based Temporal Knowledge Graph Completion</i>	
Ömer Faruk Akgül, Feiyu Zhu, Yuxin Yang, Rajgopal Kannan and Viktor Prasanna	1943
<i>Barriers to Discrete Reasoning with Transformers: A Survey Across Depth, Exactness, and Bandwidth</i>	
Michelle Yuan, Weiyi Sun, Amir H. Rezaeian, Jyotika Singh, Sandip Ghoshal, Yao-Ting Wang, Miguel Ballesteros and Yassine Benajiba	1966
<i>PaperSearchQA: Learning to Search and Reason over Scientific Papers with RLVR</i>	
James Burgess, Jan N. Hansen, Duo Peng, Yuhui Zhang, Alejandro Lozano, Min Woo Sun, Emma Lundberg and Serena Yeung-Levy	1979
<i>Too Open for Opinion? Embracing Open-Endedness in Large Language Models for Social Simulation</i>	
Bolei Ma, Yong Cao, Indira Sen, Anna-Carolina Haensch, Frauke Kreuter, Barbara Plank and Daniel Hershcovich	1998
<i>Respecting Temporal-Causal Consistency: Entity-Event Knowledge Graph for Retrieval-Augmented Generation</i>	
Ze Yu Zhang, Zitao Li, Yaliang Li, Bolin Ding and Bryan Kian Hsiang Low	2017
<i>Knowledge Extraction on Semi-Structured Content: Does It Remain Relevant for Question Answering in the Era of LLMs?</i>	
Kai Sun, Yin Huang, Srishti Mehra, Mohammad Kachuee, Xilun Chen, Renjie Tao, Zhaojiang Lin, Andrea Jessee, Nirav Shah, Alex L Betty, Yue Liu, Anuj Kumar, Wen-tau Yih and Xin Luna Dong	2055
<i>A Computational Approach to Visual Metonymy</i>	
Saptarshi Ghosh, Linfeng Liu and Tianyu Jiang	2075
<i>A Tale of Two Scripts: Transliteration and Post-Correction for Judeo-Arabic</i>	
Juan Moreno Gonzalez, Bashar Alhafni and Nizar Habash	2100
<i>Multimodal Evaluation of Russian-language Architectures</i>	
Artem Chervyakov, Ulyana Isaeva, Anton Emelyanov, Artem Safin, Maria Tikhonova, Alexander Kharitonov, Yulia Lyakh, Petr Surovtsev, Denis Shevelev, Vildan Saburov, Vasily Kononov, Elisei Rykov, Ivan Sviridov, Amina Miftakhova, Ilseyar Alimova, Alexander Panchenko, Alexander Kapitanov and Alena Fenogenova	2114

<i>Don't Judge a Book by its Cover: Testing LLMs' Robustness Under Logical Obfuscation</i>	
Abhilekh Borah, Shubhra Ghosh, Kedar Joshi, Aditya Kumar Guru and Kripabandhu Ghosh	2162
<i>I know you are different! Towards Persona Driven Knowledge-infused Dialogue Assistant</i>	
Shifali Agrahari, Moushumi Mahato, Abhisek Tiwari and Javaid Nabi	2181
<i>Enhancing Auto-regressive Chain-of-Thought through Loop-Aligned Reasoning</i>	
Qifan Yu, Zhenyu He, Sijie Li, Zhou Xun, Jun Zhang, Jingjing Xu and Di He	2206
<i>Layer-wise Swapping for Generalizable Multilingual Safety</i>	
Hyunseo Shin and Wonseok Hwang	2223
<i>Measuring Idiomaticity in Text Embedding Models with epsilon-compositionality</i>	
Sondre Wold, Étienne Simon, Erik Velldal and Lilja Øvreliid	2239
<i>Are My Optimized Prompts Compromised? Exploring Vulnerabilities of LLM-based Optimizers</i>	
Andrew Zhao, Reshmi Ghosh, Vitor R. Carvalho, Emily Lawton, Keegan Hines, Gao Huang and Jack W. Stokes	2253
<i>MAViS: A Multi-Agent Framework for Long-Sequence Video Storytelling</i>	
Qian Wang, Ziqi Huang, Ruoxi Jia, Paul Debevec and Ning Yu	2273
<i>Computational Benchmarks for Egyptian Arabic Child Directed Speech</i>	
Salam Khalifa, Abed Qaddoumi, Nizar Habash and Owen Rambow	2296
<i>K-LegalDeID: A Benchmark Dataset and KLUEBERT-CRF for De-identification in Korean Court Judgments</i>	
Wooseok Choi, Hyunbin Kim and Yon Dohn Chung	2308
<i>Specialization through Collaboration: Understanding Expert Interaction in Mixture-of-Expert Large Language Models</i>	
Yuanbo Tang, Naifan Zhang, Yan Tang, Meixuan Chen, Shuhan Huang, Tingyu Cao and Yang Li	2326
<i>Compact Language Models with Iterative Text Refinement for Health Dialogue Summarization</i>	
Kellen Tan Cheng, Ganesh Ramesh, Nafiul Rashid, Geoffrey Jay Tso and Jilong Kuang	2340
<i>Mind the Gap: Benchmarking LLM Uncertainty and Calibration with Specialty-Aware Clinical QA and Reasoning-Based Behavioural Features</i>	
Alberto Testoni and Iacer Calixto	2364
<i>Controlling Reading Ease with Gaze-Guided Text Generation</i>	
Andreas Säuberli, Darja Jepifanova, Diego Frassinelli and Barbara Plank	2383
<i>PictureStories: Predicting the Task Adherence of Language Learner Answers to a Picture Story-Based Writing Task</i>	
Marie Bexte, Andrew Caines, Diane Nicholls, Paula Buttery and Torsten Zesch	2398
<i>Assessing the Impact of Typological Features on Multilingual Machine Translation in the Age of Large Language Models</i>	
Vitalii Hirak, Jaap Jumelet and Arianna Bisazza	2416
<i>Large Language Models as Oracles for Ontology Alignment</i>	
Sviatoslav Lushnei, Dmytro Shumskyi, Severyn Shykula, Ernesto Jiménez-Ruiz and Artur d'Avila Garcez	2435

<i>Reasoning or Knowledge: Stratified Evaluation of Biomedical LLMs</i> Rahul Thapa, Qingyang Wu, Kevin Wu, Harrison G Zhang, Angela Zhang, Eric Wu, Haotian Ye and James Zou	2450
<i>Effective QA-Driven Annotation of Predicate–Argument Relations Across Languages</i> Jonathan Davidov, Aviv Slobodkin, Shmuel Tomi Klein, Reut Tsarfaty, Ido Dagan and Ayal Klein	2484
<i>Form and Meaning in Intrinsic Multilingual Evaluations</i> Wessel Poelman and Miryam de Lhoneux	2503
<i>What Breaks Knowledge Graph based RAG? Benchmarking and Empirical Insights into Reasoning under Incomplete Knowledge</i> Dongzhuoran Zhou, Yuqicheng Zhu, Xiaxia Wang, Hongkuan Zhou, Yuan He, Jiaoyan Chen, Steffen Staab and Evgeny Kharlamov	2522
<i>Assessing Web Search Credibility and Response Groundedness in Chat Assistants</i> Ivan Vykopal, Matúš Pikuliak, Simon Ostermann and Marian Simko	2539
<i>When the Model Said ‘No Comment’, We Knew Helpfulness Was Dead, Honesty Was Alive, and Safety Was Terrified</i> Gautam Siddharth Kashyap, Mark Dras and Usman Naseem	2561
<i>NeuronMoE: Efficient Cross-Lingual Extension via Neuron-Guided Mixture-of-Experts</i> Rongzhi Li and Hitomi Yanaka	2573
<i>From Emotion to Expression: Theoretical Foundations and Resources for Fear Speech</i> Vigneshwaran Shankaran, Gabriella Lapesa and Claudia Wagner	2587
<i>AdaptBPE: From General Purpose to Specialized Tokenizers</i> Vijini Pilana Liyanage and François Yvon	2607
<i>Reassessing Active Learning Adoption in Contemporary NLP: A Community Survey</i> Julia Romberg, Christopher Schröder, Julius Gonsior, Katrin Tomanek and Fredrik Olsson .	2621
<i>Beyond Not Novel Enough": Enriching Scholarly Critique with LLM-Assisted Feedback</i> Osama Mohammed Afzal, Preslav Nakov, Tom Hope and Iryna Gurevych	2648
<i>AfriVox: Probing Multilingual and Accent Robustness of Speech LLMs</i> Busayo Awobade, Mardhiyah Sanni, Tassallah Abdullahi, Chibuzor Okocha, Kelechi Ezema, Devendra Deepak Kayande, Lukman Enegi Ismaila, Tobi Olatunji and Gloria Ashiya Katuka	2672
<i>PortOldBERT: Portuguese Historical Language Models</i> Tomas Freitas Osorio and Henrique Lopes Cardoso	2691
<i>ReMedQA: Are We Done With Medical Multiple-Choice Benchmarks?</i> Alessio Cocchieri, Luca Ragazzi, Giuseppe Tagliavini and Gianluca Moro	2706
<i>Can Activation Steering Generalize Across Languages? A Study on Syllogistic Reasoning in Language Models</i> Gabriele Maraia, Leonardo Ranaldi, Marco Valentino and Fabio Massimo Zanzotto	2739
<i>SPARTA: Evaluating Reasoning Segmentation Robustness through Black-Box Adversarial Paraphrasing in Text Autoencoder Latent Space</i> Viktoriia Zinkovich, Anton Antonov, Andrei Spiridonov, Denis Shepelev, Andrey Moskalenko, Daria Pugacheva, Elena Tutubalina, Andrey Kuznetsov and Vlad Shakhuro	2754

<i>Knowledge Augmentation Enhances Token Classification for Recipe Understanding</i> Nuhu Ibrahim, Robert Stevens and Riza Batista-Navarro	2776
<i>Argumentation and Judgement Factors: LLM-based Discovery and Application in Insurance Disputes</i> Basit Ali, Anubhav Sinha, Nitin Ramrakhiyani, Sachin Pawar, Girish Keshav Palshikar and Manoj Apte	2789
<i>ViGoEmotions: A Benchmark Dataset For Fine-grained Emotion Detection on Vietnamese Texts</i> Tran Quang Hung, Pham Tien Nam, Son T. Luu and Kiet Van Nguyen	2805
<i>PTEB: Towards Robust Text Embedding Evaluation via Stochastic Paraphrasing at Evaluation Time with LLMs</i> Manuel Frank and Haithem Afli	2832
<i>DETECT: Determining Ease and Textual Clarity of German Text Simplifications</i> Maria Korobeynikova, Alessia Battisti, Lukas Fischer and Yingqiang Gao	2852
<i>MathEDU: Feedback Generation on Problem-Solving Processes for Mathematical Learning Support</i> Wei-Ling Hsu, Yu-Chien Tang and An-Zi Yen	2883
<i>Test-Time Scaling of Reasoning Models for Machine Translation</i> Zihao Li, Shaoxiong Ji and Jörg Tiedemann	2902
<i>How Good Are LLMs at Processing Tool Outputs?</i> Kiran Kate, Yara Rizk, Poulami Ghosh, Ashu Gulati, Tathagata Chakraborti, Zidane Wright and Mayank Agarwal	2918
<i>Tug-of-war between idioms' figurative and literal interpretations in LLMs</i> Soyoung Oh, Xinting Huang, Mathis Pink, Michael Hahn and Vera Demberg	2942
<i>Do LLM hallucination detectors suffer from low-resource effect?</i> Debtanu Datta, Mohan Kishore Chilukuri, Yash Kumar, Saptarshi Ghosh and Muhammad Bilal Zafar	2959
<i>Coupling Local Context and Global Semantic Prototypes via a Hierarchical Architecture for Rhetorical Roles Labeling</i> Anas Belfathi, Nicolas Hernandez, Monceaux Laura, Warren Bonnard, Mary Catherine Lavissière, Christine Jacquin and Richard Dufour	2986
<i>Guided by the Plan: Enhancing Faithful Autoregressive Text-to-Audio Generation with Guided Decoding</i> Juncheng Wang, Zhe Hu, Chao Xu, Siyue Ren, Yuxiang Feng, Yang Liu, Baigui Sun and Shujun Wang	3005
<i>Safe-Unsafe Concept Separation Emerges from a Single Direction in Language Models Activation Space</i> Andrea Ermellino, Lorenzo Malandri, Fabio Mercorio and Antonio Serino	3019
<i>PEFT-Bench: A Parameter-Efficient Fine-Tuning Methods Benchmark</i> Robert Belanec, Branislav Pecher, Ivan Srba and Maria Bielikova	3035
<i>Decoding the Market's Pulse: Context-Enriched Agentic Retrieval Augmented Generation for Predicting Post-Earnings Price Shocks</i> Chenhui Li and Weihai Lu	3055
<i>LAILA: A Large Trait-Based Dataset for Arabic Automated Essay Scoring</i> May Bashendy, Walid Massoud, Sohaila Eltanbouly, Salam Albatarni, Marwan Sayed, Abrar Abir, Houda Bouamor and Tamer Elsayed	3074

<i>Live API-Bench: 2500+ Live APIs for Testing Multi-Step Tool Calling</i>	
Benjamin Elder, Anupama Murthi, Jungkoo Kang, Ankita Naik, Kinjal Basu, Kiran Kate and Danish Contractor	3092
<i>MALicious INTent Dataset and Inoculating LLMs for Enhanced Disinformation Detection</i>	
Arkadiusz Modzelewski, Witold Sosnowski, Eleni Papadopulos, Elisa Sartori, Tiziano Labruna, Giovanni Da San Martino and Adam Wierzbicki	3125
<i>When Meanings Meet: Investigating the Emergence and Quality of Shared Concept Spaces during Multilingual Language Model Training</i>	
Felicia Körner, Max Müller-Eberstein, Anna Korhonen and Barbara Plank	3149
<i>Expanding the Boundaries of Vision Prior Knowledge in Multi-modal Large Language Models</i>	
Qiao Liang, Yanjiang Liu, Weixiang Zhou, Ben He, Yaojie Lu, Hongyu Lin, Jia Zheng, Xianpei Han, Le Sun and Yingfei Sun	3170
<i>Classifying and Addressing the Diversity of Errors in Retrieval-Augmented Generation Systems</i>	
Kin Kwan Leung, Mouloud Belbahri, Yi Sui, Alex Labach, Xueying Zhang, Stephen Anthony Rose and Jesse C. Cresswell	3185
<i>Helios: A Foundational Language Model for Smart Energy Knowledge Reasoning and Application</i>	
Haoyu Jiang, Fanjie Zeng, Boan Qu, Xiaojie Lin and Wei Zhong	3208
<i>AudioSAE: Towards Understanding of Audio-Processing Models with Sparse AutoEncoders</i>	
Georgii Aparin, Tasnima Sadekova, Alexey Rukhovich, Assel Yermekova, Laida Kushnareva, Vadim Popov, Kristian Kuznetsov and Irina Piontkovskaya	3221
<i>Vision-Language Models Align with Human Neural Representations in Concept Processing</i>	
Anna Bavaresco, Marianne De Heer Kloots, Sandro Pezzelle and Raquel Fernández	3255
<i>FAID: Fine-grained AI-generated Text Detection using Multi-task Auxiliary and Multi-level Contrastive Learning</i>	
Minh Ngoc Ta, Dong Cao Van, Duc-Anh Hoang, Minh Le-Anh, Truong Nguyen, My Anh Tran Nguyen, Yuxia Wang, Preslav Nakov and Dinh Viet Sang	3275
<i>BabyBabelLM: A Multilingual Benchmark of Developmentally Plausible Training Data</i>	
Jaap Jumelet, Abdellah Fourtassi, Akari Haga, Bastian Bunzeck, Bhargav Shandilya, Diana Galvan-Sosa, Faiz Ghifari Haznitrama, Francesca Padovani, Francois Meyer, Hai Hu, Julen Etxaniz, Laurent Prevot, Linyang He, María Grandury, Mila Marcheva, Negar Foroutan, Nikitas Theodoropoulos, Pouya Sadeghi, Siyuan Song, Suchir Salhan, Susana Zhou, Yurii Paniv, Ziyin Zhang, Arianna Bisazza, Alex Warstadt and Leshem Choshen	3297
<i>Personality Editing for Language Models through Adjusting Self-Referential Queries</i>	
Seojin Hwang, Yumin Kim, Byeongjeong Kim, Donghoon Shin and Hwanhee Lee	3330
<i>How Much Pretraining Does Structured Data Need?</i>	
Daniel Fadlon and Kfir Bar	3352
<i>Finding Culture-Sensitive Neurons in Vision-Language Models</i>	
Xiutian Zhao, Rochelle Choenni, Rohit Saxena and Ivan Titov	3366
<i>Polyglots or Multitudes? Multilingual LLM Answers to Value-laden Multiple-Choice Questions</i>	
Léo Labat, Etienne Ollion and François Yvon	3382
<i>ABCD-LINK: Annotation Bootstrapping for Cross-Document Fine-Grained Links</i>	
Serwar Basch, Iliia Kuznetsov, Tom Hope and Iryna Gurevych	3399

<i>Decision-Making with Deliberation: Meta-reviewing as a Document-grounded Dialogue</i> Sukannya Purkayastha, Nils Dycke, Anne Lauscher and Iryna Gurevych.....	3424
<i>HalluZig: Hallucination Detection using Zigzag Persistence</i> Shreyas N. Samaga, Gilberto Gonzalez Arroyo and Tamal K. Dey	3466
<i>Mapping the Course for Prompt-based Structured Prediction</i> Matt Pauk and Maria Leonor Pacheco	3483
<i>Breach in the Shield: Unveiling the Vulnerabilities of Large Language Models</i> Runpeng Dai, Run Yang, Fan Zhou and Hongtu Zhu.....	3509
<i>Martingale Foresight Sampling: A Principled Approach to Inference-Time LLM Decoding</i> Huayu Li, ZhengXiao He, Siyuan Tian, Jinghao Wen and Ao Li	3522
<i>Is This LLM Library Learning? Evaluation Must Account For Compute and Behaviour</i> Ian Berlot-Attwell, Tobias Sesterhenn, Frank Rudzicz and Xujie Si	3534
<i>Too Many Frames, Not All Useful: Efficient Strategies for Long-Form Video QA</i> Jongwoo Park, Kanchana Ranasinghe, Kumara Kahatapitiya, Wonjeong Ryu, Donghyun Kim and Michael S Ryoo.....	3569
<i>A Unified View on Emotion Representation in Large Language Models</i> Aishwarya Maheswaran and Maunendra Sankar Desarkar	3589
<i>TRACE: A Framework for Analyzing and Enhancing Stepwise Reasoning in Vision-Language Models</i> Shima Imani, Seungwhan Moon, Lambert Mathias, Lu Zhang and Babak Damavandi	3611
<i>ARC: Argument Representation and Coverage Analysis for Zero-Shot Long Document Summarization with Instruction Following LLMs</i> Mohamed Elaraby and Diane Litman	3626
<i>AudioJudge: Understanding What Works in Large Audio Model Based Speech Evaluation</i> Potsawee Manakul, Woody Haosheng Gan, Michael J Ryan, Ali Sartaz Khan, Warit Sirichote-dumrong, Kunat Pipatanakul, William Barr Held and Diyi Yang	3644
<i>Learning Multilingual Agentic Policy to Control Sycophancy</i> Leonardo Ranaldi and Giulia Pucci	3664
<i>ToxiPrompt: A Two-Stage Red-Teaming Approach for Balancing Adversarial Prompt Diversity and Response Toxicity</i> Seungho Lee and Kyumin Lee	3682
<i>AfriMTEB and AfriE5: Benchmarking and Adapting Text Embedding Models for African Languages</i> Kosei Uemura, Miaoran Zhang and David Ifeoluwa Adelani.....	3697
<i>Better Generalizing to Unseen Concepts: An Evaluation Framework and An LLM-Based Auto-Labeled Pipeline for Biomedical Concept Recognition</i> Shanshan Liu, Noriki Nishida, Fei Cheng, Narumi Tokunaga, Rumana Ferdous Munne, Yuki Yamagata, Kouji Kozaki, Takehito Utsuro and Yuji Matsumoto	3718
<i>A Representation Sharpening Framework for Zero Shot Dense Retrieval</i> Dhananjay Ashok, Suraj Nair, Mutasem Al-Darabsah, Choon Hui Teo, Tarun Agarwal and Jonathan May.....	3735
<i>Spotlight Your Instructions: Instruction-following with Dynamic Attention Steering</i> Praveen Venkateswaran and Danish Contractor.....	3752

<i>FormGym: Doing Paperwork with Agents</i>	
Matthew Toles, Isaac Song, Rattandeep Singh and Zhou Yu	3771
<i>NarraBench: A Comprehensive Framework for Narrative Benchmarking</i>	
Sil Hamilton, Matthew Wilkens and Andrew Piper	3786
<i>FaithLM: Towards Faithful Explanations for Large Language Models</i>	
Yu-Neng Chuang, Guanchu Wang, Chia-Yuan Chang, Ruixiang Tang, Shaochen Zhong, Fan Yang, Andrew Wen, Mengnan Du, Xuanting Cai, Vladimir Braverman and Xia Hu	3802
<i>Is Information Density Uniform when Utterances are Grounded on Perception and Discourse?</i>	
Matteo Gay, Coleman Haley, Mario Giulianelli and Edoardo Ponti	3825
<i>KAD: A Framework for Proxy-based Test-time Alignment with Knapsack Approximation Deferral</i>	
Ayoub Hammal, Pierre Zweigenbaum and Caio Corro	3854
<i>When Can We Trust LLMs in Mental Health? Large-Scale Benchmarks for Reliable LLM Evaluation</i>	
Abeer Badawi, Elahe Rahimi, Md Tahmid Rahman Laskar, Sheri Grach, Lindsay Bertrand, Lames Danok, Prathiba Dhanesh, Jimmy Huang, Frank Rudzicz and Elham Dolatabadi	3873
<i>DocPolarBERT: A Pre-trained Model for Document Understanding with Relative Polar Coordinate Encoding of Layout Structures</i>	
Benno Uthayasooriyar, Antoine LY, Franck Vermet and Caio Corro	3897
<i>IDEAlign: Comparing Ideas of Large Language Models to Domain Experts</i>	
HyunJi Nam, Lucía Langlois, Jim Malamut, Mei Tan and Dorottya Demszky	3908
<i>Amory: Building Coherent Narrative-Driven Agent Memory through Agentic Reasoning</i>	
Yue Zhou, Xiaobo Guo, Belhassen Bayar and Srinivasan H. Sengamedu	3926
<i>It's All About the Confidence: An Unsupervised Approach for Multilingual Historical Entity Linking using Large Language Models</i>	
Cristian Santini, Marieke van Erp and Mehwish Alam	3939
<i>SoS: Analysis of Surface over Semantics in Multilingual Text-To-Image Generation</i>	
Carolin Holtermann, Florian Schneider and Anne Lauscher	3955
<i>Gender and Politeness Perception: A Novel Approach for Exploring Annotations Disagreement</i>	
Ahmad Aljanaideh	3996
<i>TempViz: On the Evaluation of Temporal Knowledge in Text-to-Image Models</i>	
Carolin Holtermann, Nina Krebs and Anne Lauscher	4006
<i>ToxiGAN: Toxic Data Augmentation via LLM-Guided Directional Adversarial Generation</i>	
Peiran Li, Jan Fillies and Adrian Paschke	4029
<i>Text Classification Under Class Distribution Shift: A Survey</i>	
Adriana Valentina Costache, Silviu-Florin Gheorghe, Eduard Poesina, Paul Irofti and Radu Tudor Ionescu	4045
<i>Reasoning's Razor: Reasoning Improves Accuracy but Hurts Recall at Critical Operating Points in Safety and Hallucination Detection</i>	
Atoosa Chegini, Hamid Kazemi, Garrett Souza, Maria Safi, Yang Song, Samy Bengio, Sinead Williamson and Mehrdad Farajtabar	4061
<i>Instructional Agents: Reducing Teaching Faculty Workload through Multi-Agent Instructional Design</i>	
Huaiyuan Yao, Wanpeng Xu, Justin Turnau, Nadia Kellam and Hua Wei	4087

<i>Rethinking Reading Order: Toward Generalizable Document Understanding with LLM-based Relation Modeling</i>	
Weishi Wang, Hengchang Hu and Daniel Dahlmeier	4110
<i>Validating Automatic Evaluation of Controllable Counterspeech Generation: Rankings Matter More Than Scores</i>	
Yi Zheng, Björn Ross and Walid Magdy	4131
<i>Journey Before Destination: On the importance of Visual Faithfulness in Slow Thinking</i>	
Rheeya Uppaal, Phu Mon Htut, Min Bai, Nikolaos Pappas, Zheng Qi and Sandesh Swamy .	4147
<i>Automating Android Build Repair: Bridging the Reasoning-Execution Gap in LLM Agents with Domain-Specific Tools</i>	
Ha Min Son, Huan Ren, Xin Liu and Zhe Zhao	4169
<i>MetaLead: A Comprehensive Human-Curated Leaderboard Dataset for Transparent Reporting of Machine Learning Experiments</i>	
Roelien C. Timmer, Necva Bölücü and Stephen Wan	4190
<i>Enhancing the Safety of Medical Vision-Language Models by Synthetic Demonstrations</i>	
Zhiyu Xue, Reza Abbasi-Asl and Ramtin Pedarsani	4207
<i>HateXScore: A Metric Suite for Evaluating Reasoning Quality in Hate Speech Explanations</i>	
Yujia Hu and Roy Ka-Wei Lee	4221
<i>Measuring Mechanistic Independence: Can Bias Be Removed Without Erasing Demographics?</i>	
Zhengyang Shan and Aaron Mueller	4241
<i>A Survey on LLM-based Conversational User Simulation</i>	
Bo Ni, Yu Wang, Leyao Wang, Branislav Kveton, Franck Dernoncourt, Yu Xia, Hongjie Chen, Reuben Luera, Samyadeep Basu, Subhojyoti Mukherjee, Puneet Mathur, Nesreen K. Ahmed, Junda Wu, Li Li, Huixin Zhang, Ruiyi Zhang, Tong Yu, Sungchul Kim, Jiuxiang Gu, Zhengzhong Tu, Alexa Siu, Zichao Wang, Seunghyun Yoon, Nedim Lipka, Namyong Park, Zihao Lin, Trung Bui, Yue Zhao, Tyler Derr and Ryan A. Rossi	4266
<i>Prompt-driven Detection of Offensive Urdu Language using Large Language Models</i>	
Iffat Maab, Usman Haider and Junichi Yamagishi	4302
<i>Zer0-Jack: A memory-efficient gradient-based jailbreaking method for black box Multi-modal Large Language Models</i>	
Tiejun Chen, Kaishen Wang and Hua Wei	4328
<i>RAGPPI: Retrieval-Augmented Generation Benchmark for Protein-Protein Interactions in Drug Discovery</i>	
Youngseung Jeon, Ziwen Li, Thomas Li, JiaSyuan Chang, Morteza Ziyadi and Xiang Anthony Chen	4345
<i>Don't Generate, Classify! Low-Latency Prompt Optimization with Structured Complementary Prompt</i>	
Hee-Soo Kim, Jun-Young Kim, Jeong-Hwan Lee, Seong-Jin Park and Kang-Min Kim	4364
<i>CHROMIC: Chronological Reasoning Across Multi-Panel Comics</i>	
Bingxuan Hou, Jiayi Lin, Chenyang Zhang, Dapeng Yin, Shuyue Zhu, Qingqing Hong, Mengna Gao and Junli Wang	4384
<i>GAST: Gradient-aligned Sparse Tuning of Large Language Models with Data-layer Selection</i>	
Kai Yao, Zhenghan Song, Kaixin Wu, Mingjie Zhong, Danzhao Cheng, Zhaorui Tan, Yixin Ji and Penglei Gao	4401

<i>BeDiscover: The Benchmark of Discourse Understanding in the Era of Reasoning Language Models</i> Chuyuan Li and Giuseppe Carenini	4417
<i>Confidence-Calibrated Small-Large Language Model Collaboration for Cost-Efficient Reasoning</i> Chuang Zhang, Zizhen Zhu, Yihao Wei, Bing Tian, Junyi Liu, Henan Wang, Wang Xavier and Yaxiao Liu	4480
<i>Chat-Ghosting: Methods for Auto-Completion in Dialog Systems</i> Anubhab Mandal, Sandeep Mishra, Bishal Santra, Tushar Abhishek, Pawan Goyal and Manish Gupta	4502
<i>Attribution-Guided Multi-Object Hallucination and Bias Detection in Vision-Language Models</i> Sirat Samyoun, Yingtai Xiao and Jian Du	4529
<i>Word Surprisal Correlates with Sentential Contradiction in LLMs</i> Ning Shi, Bradley Hauer, David Basil, John Zhang and Grzegorz Kondrak	4549
<i>ARREST: Adversarial Resilient Regulation Enhancing Safety and Truth in Large Language Models</i> Sharanya Dasgupta, Arkaprabha Basu, Sujoy Nath and Swagatam Das	4565
<i>Re2-DocRED: Revisiting Revisited-DocRED for Joint Entity and Relation Extraction</i> Chen Kim Heng, Shao Wen Tong and Julian Wong Wei Sheng	4585
<i>Where Do LLMs Compose Meaning? A Layerwise Analysis of Compositional Robustness</i> Nura Aljaafari, Danilo Carvalho and Andre Freitas	4622
<i>BLEnD-Vis: Benchmarking Multimodal Cultural Understanding in Vision Language Models</i> Bryan Chen Zhengyu Tan, Weihua Zheng, Zhengyuan Liu, Nancy F. Chen, Hwaran Lee, Kenny Tsu Wei Choo and Roy Ka-Wei Lee	4647
<i>Document-Level Zero-Shot Relation Extraction with Entity Side Information</i> Mohan Raj, Lay-Ki Soon, Huey Fang Ong and Bhawani Selvaretnam	4670
<i>Steering Large Language Models for Machine Translation Personalization</i> Daniel Scalena, Gabriele Sarti, Arianna Bisazza, Elisabetta Fersini and Malvina Nissim	4681
<i>Taxation Perspectives from Large Language Models: A Case Study on Additional Tax Penalties</i> Eunkyung Choi, Young Jin Suh, Siun Lee, Hongseok Oh, Juheon Kang, Won Hur, Hun Park and Wonseok Hwang	4702
<i>Beyond Memorization: A Rigorous Evaluation Framework for Medical Knowledge Editing</i> Shigeng Chen, Linhao Luo, Zhangchi Qiu, Yanan Cao, Carl Yang and Shirui Pan	4727
<i>Unlocking Latent Discourse Translation in LLMs Through Quality-Aware Decoding</i> Wafaa Mohammed, Vlad Niculae and Chrysoula Zerva	4752
<i>Cross-lingual and Word-Independent Methods for Quantifying Degree of Grammaticalization</i> Ryo Nagata, Daichi Mochihashi, Misato Ido, Yusuke Kubota, Naoki Otani, Yoshifumi Kawasaki and Hiroya Takamura	4775
<i>Knowing the Facts but Choosing the Shortcut: Understanding How Large Language Models Compare Entities</i> Hans Hergen Lehmann, Jae Hee Lee, Steven Schockaert and Stefan Wermter	4788
<i>Calibrating Beyond English: Language Diversity for Better Quantized Multilingual LLMs</i> Everlyn Asiko Chimoto, Mostafa Elhoushi and Bruce Bassett	4822

<i>LaCoMSA: Language-Consistency Multilingual Self-Alignment with Latent Representation Rewarding</i> Khanh-Tung Tran, Barry O’Sullivan and Hoang D. Nguyen	4839
<i>Can you map it to English? The Role of Cross-Lingual Alignment in the Multilingual Performance of LLMs</i> Kartik Ravisankar, HyoJung Han, Sarah Wiegrefe and Marine Carpuat	4854
<i>Recursive numeral systems are highly regular and easy to process</i> Ponrawee Prasertsom, Andrea Silvi, Jennifer Culbertson, Devdatt Dubhashi, Moa Johansson and Kenny Smith	4873
<i>Bringing Emerging Architectures to Sequence Labeling in NLP</i> Ana Ezquerro, Carlos Gómez-Rodríguez and David Vilares	4886
<i>SEMIROUTER: Sparse-Data Enhanced Routing for Adaptive Multi-LLM System</i> Zijie Wang, Xinyu Yan, Che Wang, Zeng Zihao, Lei Xiao and Wei Yang Bryan Lim	4910
<i>DITTO: A Spoofing Attack Framework on Watermarked LLMs via Knowledge Distillation</i> Hyeseon An, Shinwoo Park, Suyeon Woo and Yo-Sub Han	4922
<i>Boundary-Aware LLM Augmentation for Low-Resource Event Argument Extraction</i> Zhaoyue Sun, Gabriele Pergola and Yulan He	4937
<i>CASE – Condition-Aware Sentence Embeddings for Conditional Semantic Textual Similarity Measurement</i> Gaifan Zhang, Yi Zhou and Danushka Bollegala	4954
<i>Evaluation and LLM-Guided Learning of ICD Coding Rationales</i> Mingyang Li, Viktor Schlegel, Tingting Mu, Wuraola Oyewusi, Kai Kang and Goran Nenadic	4969
<i>Evaluating the Effect of Retrieval Augmentation on Social Biases</i> Tianhui Zhang, Yi Zhou and Danushka Bollegala	5004
<i>Persuasion at Play: Understanding Misinformation Dynamics in Demographic-Aware Human-LLM Interactions</i> Angana Borah, Rada Mihalcea and Veronica Perez-Rosas	5027
<i>Entropy-Gated Branching for Efficient Test-Time Reasoning</i> Xianzhi Li, Ethan Callanan, Abdellah Ghassel and Xiaodan Zhu	5054
<i>Decomposition-Enhanced Training for Post-Hoc Attributions in Language Models</i> Sriram Balasubramanian, Samyadeep Basu, Koustava Goswami, Ryan A. Rossi, Varun Manjunatha, Roshan Santhosh, Ruiyi Zhang, Soheil Feizi and Nedim Lipka	5070
<i>INSURE-Dial: A Phase-Aware Conversational Dataset & Benchmark for Compliance Verification and Phase Detection</i> Shubham Kulkarni, Alexander Lyzhov, Preetam Joshi and Shiva Chaitanya	5085
<i>NLP for Social Good: A Survey and Outlook of Challenges, Opportunities and Responsible Deployment</i> Antonia Karamolegkou, Angana Borah, Eunjung Cho, Sagnik Ray Choudhury, Martina Galletti, Pranav Gupta, Oana Ignat, Priyanka Kargupta, Neema Kotonya, Hemank Lamba, Sun-Joo Lee, Arushi Mangla, Ishani Mondal, Fatima Zahra Moudakir, Deniz Nazar, Poli Nemkova, Dina Pisarevskaya, Naqee Rizwan, Nazanin Sabri, Keenan Samway, Dominik Stambach, Anna Steinberg Schulten, David Tomás, Steven R Wilson, Bowen Yi, Jessica H Zhu, Arkaitz Zubiaga, Anders Sjøgaard, Alexander Fraser, Zhijing Jin, Rada Mihalcea, Joel R. Tetreault and Daryna Dementieva	5110

<i>From Delegates to Trustees: How Optimizing for Long-Term Interests Shapes Bias and Alignment in LLMs</i>	Suyash Fulay, Jocelyn Zhu and Michiel A. Bakker	5171
<i>Investigating Language and Retrieval Bias in Multilingual Previously Fact-Checked Claim Detection</i>	Ivan Vykopal, Antonia Karamolegkou, Jaroslav Kopčan, Qiwei Peng, Tomáš Javůrek, Michal Gregor and Marian Simko	5195
<i>FFE-Hallu: Hallucinations in Fixed Figurative Expressions: A Benchmark of Idioms and Proverbs in the Persian Language</i>	Faezeh Hosseini, Mohammadali Yousefzadeh and Yadollah Yaghoobzadeh	5222
<i>MEVER: Multi-Modal and Explainable Claim Verification with Graph-based Evidence Retrieval</i>	Delvin Ce Zhang, Suhan Cui, Zhelin Chu, Xianren Zhang and Dongwon Lee	5236
<i>DuwatBench: Bridging Language and Visual Heritage through an Arabic Calligraphy Benchmark for Multimodal Understanding</i>	Shubham Patle, Sara Ghaboura, Hania Tariq, Mohammad Usman Khan, Omkar Thawakar, Rao Muhammad Anwer and Salman Khan	5256
<i>ConvApparel: A Benchmark Dataset and Validation Framework for User Simulators in Conversational Recommenders</i>	Ofer Meshi, Krisztian Balog, Sally Goldman, Avi Caciularu, Guy Tennenholtz, Jihwan Jeong, Amir Globerson and Craig Boutilier	5270
<i>Detecting Latin in Historical Books with Large Language Models: A Multimodal Benchmark</i>	Yu Wu, Ke Shu, Jonas Fischer, Lidia Pivovarova, David Rosson, Eetu Mäkelä and Mikko Tolonen	5305
<i>Persistent Personas? Role-Playing, Instruction Following, and Safety in Extended Interactions</i>	Pedro Henrique Luz de Araujo, Michael A. Hedderich, Ali Modarressi, Hinrich Schuetze and Benjamin Roth	5329
<i>CliniBench: A Clinical Outcome Prediction Benchmark for Generative and Encoder-Based Language Models</i>	Paul Grundmann, Jan Frick, Dennis Fast, Thomas Steffek, Felix Gers, Wolfgang Nejdl and Alexander Löser	5360
<i>DIVINE : Coordinating Multimodal Disentangled Representations for Oro-Facial Neurological Disorder Assessment</i>	Mohd Mujtaba Akhtar, Girish and Muskaan Singh	5379
<i>Biasless Language Models Learn Unnaturally: How LLMs Fail to Distinguish the Possible from the Impossible</i>	Imry Ziv, Nur Lan and Emmanuel Chemla	5393
<i>Bridging Attribution and Open-Set Detection using Graph-Augmented Instance Learning in Synthetic Speech</i>	Mohd Mujtaba Akhtar, Girish, Farhan Sheth and Muskaan Singh	5404
<i>Detecting Non-Membership in LLM Training Data via Rank Correlations</i>	Pranav Shetty, Mirazul Haque, Zhiqiang Ma and Xiaomo Liu	5414
<i>Taming Object Hallucinations with Verified Atomic Confidence Estimation</i>	Jiarui Liu, Weihao Xuan, Zhijing Jin and Mona T. Diab	5430

<i>DART: Leveraging Multi-Agent Disagreement for Tool Recruitment in Multimodal Reasoning</i> Nithin Sivakumaran, Justin Chen, David Wan, Yue Zhang, Jaehong Yoon, Elias Stengel-Eskin and Mohit Bansal	5445
<i>ToolDreamer: Instilling LLM Reasoning Into Tool Retrievers</i> Saptarshi Sengupta, Zhengyu Zhou, Jun Araki, Xingbo Wang, Bingqing Wang, Suhang Wang and Zhe Feng	5465
<i>An Empirical Study of Speculative Decoding for Small Language Models</i> Luca Mainardi, Selcuk Sandikci and Joaquin Vanschoren	5483
<i>Lost in Formatting: How Output Formats Skew LLM Performance on Information Extraction</i> Rishi Ravikumar, Nuhu Ibrahim and Riza Batista-Navarro	5498
<i>Pseudo-Likelihood Training for Reasoning Diffusion Language Models</i> Shiv Shankar	5514
<i>RoSE: Round-robin Synthetic Data Evaluation for Selecting LLM Generators without Human Test Sets</i> Jan Cegin, Branislav Pecher, Ivan Srba and Jakub Simko	5530
<i>RotBench: Evaluating Multi-modal Large Language Models on Identifying Image Rotation</i> Tianyi Niu, Jaemin Cho, Elias Stengel-Eskin and Mohit Bansal	5546
<i>Multilingual Amnesia: On the Transferability of Unlearning in Multilingual LLMs</i> Alireza Dehghanpour Farashah, Aditi Khandelwal, Marylou Fauchard, Zhuan Shi, Negar Rostamzadeh and Golnoosh Farnadi	5570
<i>Beyond Math: Stories as a Testbed for Memorization-Constrained Reasoning in LLMs</i> Yuxuan Jiang and Francis Ferraro	5590
<i>Neural Breadcrumbs: Membership Inference Attacks on LLMs Through Hidden State and Attention Pattern Analysis</i> Disha Makhija, Manoj Ghuhana Arivazhagan, Vinayshekhar Bannihatti Kumar and Rashmi Gangadharaiah	5608
<i>Chat-TS: Enhancing Multi-Modal Reasoning Over Time-Series and Natural Language Data</i> Paul Quinlan, Qingguo Li and Xiaodan Zhu	5621
<i>Beyond Names: How Grammatical Gender Markers Bias LLM-based Educational Recommendations</i> Luca Benedetto, Antonia Donvito, Alberto Lucchetti, Andrea Cappelli and Paula Buttery ..	5648
<i>ExStrucTiny: A Benchmark for Schema-Variable Structured Information Extraction from Document Images</i> Mathieu Sibue, Andrés Muñoz Garza, Samuel Mensah, Pranav Shetty, Zhiqiang Ma, Xiaomo Liu and Manuela Veloso	5669
<i>What's Missing in Vision-Language Models? Probing Their Struggles with Causal Order Reasoning</i> Zhaotian Weng, Haoxuan Li, Xin Eric Wang, Kuan-Hao Huang and Jieyu Zhao	5689
<i>KidsArtBench: Multi-Dimensional Children's Art Evaluation with Attribute-Aware MLLMs</i> Mingrui Ye, Chanjin Zheng, Zengyi Yu, Chenyu Xiang, Zhixue Zhao, Zheng Yuan and Helen Yannakoudakis	5702
<i>Steering Safely or Off a Cliff? Rethinking Specificity and Robustness in Inference-Time Interventions</i> Navita Goyal and Hal Daumé Iii	5723

<i>Tracing Multilingual Knowledge Acquisition Dynamics in Domain Adaptation: A Case Study of Bio-medical Adaptation</i>	
Xin Zhao, Naoki Yoshinaga, Yuma Tsuta and Akiko Aizawa	5739
<i>Contextual morphologically-guided tokenization for Latin encoder models</i>	
Marisa Hudspeth, Patrick J. Burns and Brendan O’Connor	5761
<i>Beyond Random Sampling: Efficient Language Model Pretraining via Curriculum Learning</i>	
Yang Zhang, Amr Mohamed, Hadi Abdine, Guokan Shang and Michalis Vazirgiannis	5776
<i>ObjChangeVR: Object State Change Reasoning from Continuous Egocentric Views in VR Environments</i>	
Shiyi Ding, Shaoen WU and Ying Chen	5795
<i>Tracking the Limits of Knowledge Propagation: How LLMs Fail at Multi-Step Reasoning with Conflicting Knowledge</i>	
Yiyang Feng, Zeming Chen, Haotian Wu, Jiawei Zhou and Antoine Bosselut	5813
<i>Do Audio LLMs Really LISTEN, or Just Transcribe? Measuring Lexical vs. Acoustic Emotion Cues Reliance</i>	
Jingyi Chen, Zhimeng Guo, Jiyun Chun, Pichao Wang, Andrew Perrault and Micha Elsner .	5848
<i>CSPB: Conversational Speech Processing Benchmark for Self-supervised Speech Models</i>	
Zili Huang, Matthew Maciejewski, Leibny Paola Garcia Perera, Shinji Watanabe and Sanjeev Khudanpur	5878
<i>Multi-Token Completion for Text Anonymization</i>	
Pulkit Madaan, Krithika Ramesh, Lisa Bauer, Charith Peris and Anjalie Field	5894
<i>MERLIN: Multi-Stage Curriculum Alignment for Multilingual Encoder-LLM Integration in Cross-Lingual Reasoning</i>	
Kosei Uemura, David Guzmán, Quang Phuoc Nguyen, Jesujoba Oluwadara Alabi, En-Shiun Annie Lee and David Ifeoluwa Adelani	5909
<i>Now You Hear Me: Audio Narrative Attacks Against Large Audio–Language Models</i>	
Ye Yu, Haibo Jin, Yaoning Yu, Jun Zhuang and Haohan Wang	5925
<i>Evaluating Adversarial Robustness of Concept Representations in Sparse Autoencoders</i>	
Aaron J. Li, Suraj Srinivas, Usha Bhalla and Himabindu Lakkaraju	5940
<i>Mary, the Cheeseburger-Eating Vegetarian: Do LLMs Recognize Incoherence in Narratives?</i>	
Karin De Langis, Püren Öncel, Ryan Peters, Andrew Elfenbein, Laura Kristen Allen, Andreas Schramm and Dongyeop Kang	5958
<i>Strong Memory, Weak Control: An Empirical Study of Executive Functioning in LLMs</i>	
Karin De Langis, Jong Inn Park, Khanh Chi Le, Andreas Schramm, Andrew Elfenbein, Michael C. Mensink and Dongyeop Kang	5971
<i>How Do Language Models Acquire Character-Level Information?</i>	
Soma Sato and Ryohei Sasano	5987
<i>Analysing the role of lexical and temporal information in turn-taking through predictability</i>	
Sean Leishman, Sarenne Wallbridge and Peter Bell	5998
<i>Beyond Length: Context-Aware Expansion and Independence as Developmentally Sensitive Evaluation in Child Utterances</i>	
Jiyun Chun, Eric Fosler-Lussier, Michael White and Andrew Perrault	6010

<i>Translation via Annotation: A Computational Study of Translating Classical Chinese into Japanese</i> Zilong Li and Jie Cao	6031
<i>Extending Audio Context for Long-Form Understanding in Large Audio-Language Models</i> Yuatyong Chaichana, Pittawat Taveekitworachai, Warit Sirichotedumrong, Potsawee Manakul and Kunat Pipatanakul	6046
<i>HALP: Detecting Hallucinations in Vision-Language Models without Generating a Single Token</i> Sai Akhil Kogilathota, Sripadha Vallabha E G, Luzhe Sun and Jiawei Zhou	6067
<i>Nanda Family: Open-Weights Generative Large Language Models for Hindi</i> Aaryamonvikram Singh, Debopriyo Banerjee, Dhruv Sahnan, Monojit Choudhury, Shivam Chauhan, Rocktim Jyoti Das, Xudong Han, Haonan Li, Alok Anil Jadhav, Utkarsh Agarwal, Mukund Choudhary, Fajri Koto, Junaid Hamid Bhat, Awantika Shukla, Samujjwal Ghosh, Samta Kamboj, Onkar Pandit, Lalit Pradhan, Rahul Pal, Sunil Kumar Sahu, Parvez Mullah, Ali El Filali, Zainul Abedien Ahmed Quraishi, Neha Sengupta, Gokulakrishnan Ramakrishnan, Rituraj Joshi, Gurpreet Gosal, Avraham Sheinin, Natalia Vassilieva and Preslav Nakov	6086
<i>Wugnectives: Novel Entity Inferences of Language Models from Discourse Connectives</i> Daniel Brubaker, William Sheffield, Junyi Jessy Li and Kanishka Misra	6109
<i>Can LLMs reason over extended multilingual contexts? Towards long-context evaluation beyond retrieval over haystacks</i> Amey Hengle, Prasoon Bajpai, Soham Dan and Tanmoy Chakraborty	6128
<i>Knowing When to Abstain: Medical LLMs Under Clinical Uncertainty</i> Sravanthi Machcha, Sushrita Yerra, Sahil Gupta, Aishwarya Sahoo, Sharmin Sultana, Hong yu and Zonghai Yao	6153
<i>MedQA-CS: Objective Structured Clinical Examination (OSCE)-Style Benchmark for Evaluating LLM Clinical Skills</i> Zonghai Yao, Zihao Zhang, Chaolong Tang, Xingyu Bian, Youxia Zhao, Zhichao Yang, Junda Wang, Huixue Zhou, Won Seok Jang, Feiyun Ouyang and Hong yu	6183
<i>Continual-learning for Modelling Low-Resource Languages from Large Language Models</i> Santosh Srinath K, Mudit Somani, Varun Reddy Padala, Prajna Upadhyay and Abhijit Das .	6258
<i>Language-Grounded Multi-Domain Image Translation via Semantic Difference Guidance</i> Jongwon Ryu, Joonhyung Park, Jaeho Han, Yeong-Seok Kim, Hye-Rin Kim, Sunjae Yoon and Junyeong Kim	6276
<i>LLMs as Cultural Archives: Cultural Commonsense Knowledge Graph Extraction</i> Junior Cedric Tonga, Chen Cecilia Liu, Iryna Gurevych and Fajri Koto	6289
<i>Nahw: A Comprehensive Benchmark of Arabic Grammar Understanding, Error Detection, Correction, and Explanation</i> Hamdy Mubarak, Majd Hawasly and Abubakr Mohamed	6310
<i>Rethinking Creativity Evaluation: A Critical Analysis of Existing Creativity Evaluations</i> Li-Chun Lu, Miri Liu, Pin Chun Lu, Yufei Tian, Shao-Hua Sun and Nanyun Peng	6329
<i>TReX: Tokenizer Regression for Optimal Data Mixture</i> Inho Won, Hangeeol Yoo, Minkyung Cho, Jungyeul Park, Hoyun Song and KyungTae Lim	6353
<i>CONGRAD: Conflicting Gradient Filtering for Multilingual Preference Alignment</i> Jiangnan Li, Thuy-Trang Vu, Christian Herold, Amirhossein Tebbifakhr, Shahram Khadivi and Gholamreza Haffari	6371

<i>Activation-Space Personality Steering: Hybrid Layer Selection for Stable Trait Control in LLMs</i> Pranav Bhandari, Nicolas Fay, Sanjeevan Selvaganapathy, Amitava Datta, Usman Naseem and Mehwish Nasim	6388
<i>Speculative Decoding Speed-of-Light: Optimal Lower Bounds via Branching Random Walks</i> Sergey Pankratov and Dan Alistarh	6404
<i>KG-CRAFT: Knowledge Graph-based Contrastive Reasoning with LLMs for Enhancing Automated Fact-checking</i> Vítor Lourenço, Aline Paes, Tillman Weyde, Audrey Depeige and Mohnish Dubey	6419
<i>SciRAG: Adaptive, Citation-Aware, and Outline-Guided Retrieval and Synthesis for Scientific Literature</i> Hang Ding, Yilun Zhao, Tiansheng Hu, Manasi Patwardhan and Arman Cohan	6440
<i>Unintended Memorization of Sensitive Information in Fine-Tuned Language Models</i> Marton Szep, Jorge Marin Ruiz, Georgios Kaissis, Paulina Seidl, Rüdiger von Eisenhart-Rothe, Florian Hinterwimmer and Daniel Rueckert	6461
<i>The Pluralistic Moral Gap: Understanding Moral Judgment and Value Differences between Humans and Large Language Models</i> Giuseppe Russo, Debora Nozza, Paul Röttger and Dirk Hovy	6481
<i>CoReTab: Improving Multimodal Table Understanding with Code-driven Reasoning</i> Van-Quang Nguyen and Takayuki Okatani	6498
<i>Explaining Generalization of AI-Generated Text Detectors Through Linguistic Analysis</i> Yuxi Xia, Kinga Stańczak and Benjamin Roth	6524
<i>Elections go bananas: A First Large-scale Multilingual Study of Pluralia Tantum using LLMs</i> Elena Spaziani, Kamyar Zeinalipour, Pierluigi Cassotti and Nina Tahmasebi	6547
<i>CacheNotes: Task-Aware Key-Value Cache Compression for Reasoning-Intensive Knowledge Tasks</i> Giulio Corallo, Orion Weller, Fabio Petroni and Paolo Papotti	6571
<i>Beyond Blind Following: Evaluating Robustness of LLM Agents under Imperfect Guidance</i> Yao Fu, Ran Qiu, Xinhe Wang, Jacob Sansom, Sathvika Ayyappa Prabhu, Huijie Tang, Jaekyeom Kim, Sungryull Sohn and Honglak Lee	6591
<i>How Do LLMs Generate Contrastive Sentiments? A Mechanistic Perspective</i> Van Bach Nguyen, Jörg Schlötterer and Christin Seifert	6619
<i>Continual Neural Topic Model</i> Charu Karakkaparambil James, Waleed Mustafa, Marcio Monteiro, Marius Kloft and Sophie Fellenz	6636
<i>MAQuA: Multi-outcome Adaptive Question-Asking for Mental Health using Item Response Theory</i> Vasudha Varadarajan, Hui Xu, Rebecca Astrid Böhme, Mariam Marlen Mirström, Sverker Sikström and H. Andrew Schwartz	6659
<i>Principled Self-Correction in Discrete Diffusion: A UCB-Guided Framework for Text Generation</i> Masaki Asada and Makoto Miwa	6678
<i>ConLID: Supervised Contrastive Learning for Low-Resource Language Identification</i> Negar Foroutan, Jakhongir Saydaliev, Grace Kim and Antoine Bosselut	6693
<i>Emotionally Charged, Logically Blurred: AI-driven Emotional Framing Impairs Human Fallacy Detection</i> Yanran Chen, Lynn Greschner, Roman Klinger, Michael Klenk and Steffen Eger	6709

<i>Aligning Text, Code, and Vision: A Multi-Objective Reinforcement Learning Framework for Text-to-Visualization</i>	
Mizanur Rahman, Mohammed Saidul Islam, Md Tahmid Rahman Laskar, Shafiq Joty and Enamul Hoque	6733
<i>Offline Preference Optimization via Maximum Marginal Likelihood Estimation</i>	
Saeed Najafi and Alona Fyshe	6751
<i>The Relevance of Value Systems for Offensive Language Detection</i>	
Michael Wiegand, Elisabeth Eder and Josef Ruppenhofer	6765
<i>Instruction Tuning with and without Context: Behavioral Shifts and Downstream Impact</i>	
Hyunji Lee, Seunghyun Yoon, Yunjae Won, Hanseok Oh, Geewook Kim, Trung Bui, Franck Dernoncourt, Elias Stengel-Eskin, Mohit Bansal and Minjoon Seo	6790
<i>RefusalBench: Generative Evaluation of Selective Refusal in Grounded Language Models</i>	
Aashiq Muhamed, Leonardo F. R. Ribeiro, Markus Dreyer, Virginia Smith and Mona T. Diab	6811
<i>Query Decomposition for RAG: Balancing Exploration-Exploitation</i>	
Roxana Petcu, Kenton Murray, Daniel Khashabi, Evangelos Kanoulas, Maarten de Rijke, Dawn Lawrie and Kevin Duh	6857
<i>Do Images Speak Louder than Words? Investigating the Effect of Textual Misinformation in VLMs</i>	
Chi Zhang, Wenxuan Ding, Jiale Liu, Mingrui Wu, Qingyun Wu and Ray Mooney	6872
<i>Sycophancy Hides Linearly in the Attention Heads</i>	
Rifo Ahmad Genadi, Munachiso Samuel Nwadike, Nurdaulet Mukhituly, Tatsuya Hiraoka, Hilal AlQuabeh and Kentaro Inui	6896
<i>AICD Bench: A Challenging Benchmark for AI-Generated Code Detection</i>	
Daniil Orel, Dilshod Azizov, Indraneil Paul, Yuxia Wang, Iryna Gurevych and Preslav Nakov	6913
<i>Safeguarding Language Models via Self-Destruct Trapdoor</i>	
Shahar Katz, Bar Alon, Ariel Shaulov, Lior Wolf and Mahmood Sharif	6939
<i>Rethinking Hallucinations: Correctness, Consistency, and Prompt Multiplicity</i>	
Prakhar Ganesh, Reza Shokri and Golnoosh Farnadi	6959
<i>Hype or not? Formalizing Automatic Promotional Language Detection in Biomedical Research</i>	
Bojan Batalo, Erica K. Shimomoto, Dipesh Satav and Neil Millar	6979
<i>H3Fusion: Helpful, Harmless, Honest Fusion of Aligned LLMs</i>	
Selim Furkan Tekin, Fatih Ilhan, Sihao Hu, Tiansheng Huang, Yichang Xu, Zachary Yahn and Ling Liu	6993
<i>Revisiting Generalization Across Difficulty Levels: It's Not So Easy</i>	
Yeganeh Kordi, Nihal V. Nayak, Max Zuo, Ilana Nguyen and Stephen Bach	7014
<i>BLUR: A Bi-Level Optimization Approach for LLM Unlearning</i>	
Hadi Reisizadeh, Jinghan Jia, Zhiqi Bu, Bhanukiran Vinzamuri, Anil Ramakrishna, Kai-Wei Chang, Volkan Cevher, Sijia Liu and Mingyi Hong	7043
<i>DeepInsert: Early Layer Bypass for Efficient and Performant Multimodal Understanding</i>	
Moulik Choraria, Xinbo Wu, Akhil Bhimaraju, Nitesh Sekhar, Yue Wu, Xu Zhang, Prateek Singhal and Lav R. Varshney	7059

<i>Dynamic Cheatsheet: Test-Time Learning with Adaptive Memory</i>	
Mirac Suzgun, Mert Yuksekgonul, Federico Bianchi, Dan Jurafsky and James Zou	7080
<i>Evidential Semantic Entropy for LLM Uncertainty Quantification</i>	
Lucie Kunitomo-Jacquín, Edison Marrese-Taylor, Ken Fukuda and Masahiro Hamasaki	7107
<i>SCENEBench: An Audio Understanding Benchmark Grounded in Assistive and Industrial Use Cases</i>	
Laya Iyer, Angelina Wang and Sanmi Koyejo	7123
<i>Incentivizing Strong Reasoning from Weak Supervision</i>	
Yige Yuan, Teng Xiao, Shuchang Tao, Xue Wang, Jinyang Gao, Bolin Ding and Bingbing Xu	7138
<i>DivMerge: A divergence-based model merging method for multi-tasking</i>	
Brahim Touayouch, Loïc Fosse, Géraldine Damnati and Gwénoél Lecorvé	7157
<i>A Reinforcement Learning Framework for Robust and Secure LLM Watermarking</i>	
Li An, Yujian Liu, Yepeng Liu, Yuheng Bu, Yang Zhang and Shiyu Chang	7181
<i>Agent-Testing Agent: A Meta-Agent for Automated Testing and Evaluation of Conversational AI Agents</i>	
Sameer Komoravolu and Khalil Mrini	7199
<i>User-Centric Evidence Ranking for Attribution and Fact Verification</i>	
Guy Alt, Eran Hirsch, Serwar Basch, Ido Dagan and Oren Glickman	7215
<i>Beyond Understanding: Evaluating the Pragmatic Gap in LLMs' Cultural Processing of Figurative Language</i>	
Mena Attia, Aashiq Muhamed, Mai Alkhamissi, Thamar Solorio and Mona T. Diab	7238
<i>VietMix: A Naturally-Occurring Parallel Corpus and Augmentation Framework for Vietnamese-English Code-Mixed Machine Translation</i>	
Hieu Tran, Phuong-Anh Nguyen-Le, Huy Nghiem, Quang-Nhan Nguyen, Wei Ai and Marine Carpuat	7266
<i>Do You See Me : A Multidimensional Benchmark for Evaluating Visual Perception in Multimodal LLMs</i>	
Aditya Sanjiv Kanade and Tanuja Ganu	7285
<i>An Empirical Study of Collective Behaviors and Social Dynamics in Large Language Model Agents</i>	
Farnoosh Hashemi and Michael Macy	7327
<i>Detecting Subtle Biases: An Ethical Lens on Underexplored Areas in AI Language Models Biases</i>	
Shayan Bali, Farhan Farsi, Mohammad Hosseini, Adel Khorramrouz and Ehsaneddin Asgari	7352
<i>HarfoSokhan: A Comprehensive Parallel Dataset for Transitions between Persian Colloquial and Formal Variations</i>	
Hamid Jahad Sarvestani, Vida Ramezani, Saeed Saadat, Neda Taghizadeh Serajeh, Maryam Saadat Razavi Taheri, Shohreh Kasaei, Mohammad Amin Fazli and Ehsaneddin Asgari	7380
<i>Compressing Language Models for Specialized Domains</i>	
Miles Williams, George Chrysostomou, Vitor Amancio Jeronymo and Nikolaos Aletras	7393
<i>GRAVITY: A Framework for Personalized Text Generation via Profile-Grounded Synthetic Preferences</i>	
Priyanka Dey, Daniele Rosa, Wenqing Zheng, Daniel Barcklow, Jieyu Zhao and Emilio Ferrara	7416
<i>Multimodal Conversation Structure Understanding</i>	
Kent K. Chang, Mackenzie Hanh Cramer, Anna Ho, Ti Ti Nguyen, Yilin Yuan and David Bamman	7437

<i>A Review of Incorporating Psychological Theories in LLMs</i>	
Zizhou Liu, Ziwei Gong, Lin Ai, Zheng Hui, Run Chen, Colin Wayne Leach, Michelle R. Greene and Julia Hirschberg	7459
<i>How Robust Are Router-LLMs? Analysis of the Fragility of LLM Routing Capabilities</i>	
Aly M. Kassem, Bernhard Schölkopf and Zhijing Jin	7496
<i>NG-Router: Graph-Supervised Multi-Agent Collaboration for Nutrition Question Answering</i>	
Kaiwen Shi, Zheyuan Zhang, Zhengqing Yuan, Keerthiram Murugesan, Vincent Galassi, Chuxu Zhang and Yanfang Ye	7508
<i>Verification-Aware Planning for Multi-Agent Systems</i>	
Tianyang Xu, Dan Zhang, Kushan Mitra and Estevam Hruschka	7528
<i>Zero-Shot Open-Schema Entity Structure Discovery</i>	
Xueqiang Xu, Jinfeng Xiao, James Barry, Mohab Elkaref, Jiaru Zou, Pengcheng Jiang, Yunyi Zhang, Maxwell J Giammona, Geeth De Mel and Jiawei Han	7547
<i>Beyond Semantics: How Temporal Biases Shapes Retrieval in Transformer and State-Space Models</i>	
Anooshka Bajaj, Deven Mahesh Mistry, Sahaj Singh Maini, Yash Aggarwal and Zoran Tiganj	7562
<i>Diagnosing Vision Language Models’ Perception by Leveraging Human Methods for Color Vision Deficiencies</i>	
Kazuki Hayashi, Shintaro Ozaki, Yusuke Sakai, Hidetaka Kamigaito and Taro Watanabe	7582
<i>Tokenizer-Aware Cross-Lingual Adaptation of Decoder-Only LLMs through Embedding Relearning and Swapping</i>	
Fan Jiang, Honglin Yu, Grace Y Chung and Trevor Cohn	7606
<i>Active Generalized Category Discovery with Diverse LLM Feedback</i>	
Henry Peng Zou, Siffi Singh, Yi Nian, Jianfeng He, Jason Cai, Saab Mansour and Hang Su	7637
<i>RAFFLES: Reasoning-based Attribution of Faults for LLM Systems</i>	
Chenyang Zhu, Spencer Hong, Jingyu Wu, Kushal Chawla, Yuhui Tang, Youbing Yin, Nathan Wolfe, Erin Babinsky and Daben Liu	7659
<i>Jailbreaks as Inference-Time Alignment: A Framework for Understanding Safety Failures in LLMs</i>	
James Beetham, Souradip Chakraborty, Mengdi Wang, Furong Huang, Amrit Singh Bedi and Mubarak Shah	7689
<i>Over-Searching in Retrieval-Augmented Large Language Models</i>	
Roy Xie, Deepak Gopinath, David Qiu, Dong Lin, Haitian Sun, Saloni Potdar and Bhuwan Dhingra	7714
<i>LitBench: A Benchmark and Dataset for Reliable Evaluation of Creative Writing</i>	
Daniel Fein, Sebastian Russo, Violet Xiang, Kabir Jolly, Rafael Rafailov and Nick Haber	7740
<i>H-Mem: Hybrid Multi-Dimensional Memory Management for Long-Context Conversational Agents</i>	
Zihe Ye, Jingyuan Huang, Weixin Chen and Yongfeng Zhang	7756
<i>“Yuki Gets Sushi, David Gets Steak?”: Uncovering Gender and Racial Biases in LLM-Based Meal Recommendations</i>	
Xuefeng Wei, Xuan Zhou, Yusuke Sakai and Taro Watanabe	7776
<i>Happiness is Sharing a Vocabulary: A Study of Transliteration Methods</i>	
Haeji Jung, Jinju Kim, Kyungjin Kim, Youjeong Roh and David R. Mortensen	7797

<i>SCALAR: Scientific Citation-based Live Assessment of Long-context Academic Reasoning</i> Renxi Wang, Honglin Mu, Liqun Ma, Lizhi Lin, Yunlong Feng, Timothy Baldwin, Xudong Han and Haonan Li	7817
<i>Look Before You Leap: A Lookahead Reasoning Quality Gate for Speculative Decoding</i> Hiroaki Kingetsu, Kaoru Yokoo, Kenji Fukumizu and Manohar Kaul	7831
<i>FanarGuard: A Culturally-Aware Moderation Filter for Arabic Language Models</i> Masoomali Fatehkia, Enes Altinisik and Husrev Taha Sencar	7848
<i>BILLY: Steering Large Language Models via Merging Persona Vectors for Creative Generation</i> Tsung-Min Pai, Jui-I Wang, Li-Chun Lu, Shao-Hua Sun, Hung-yi Lee and Kai-Wei Chang .	7870
<i>Unveiling Intrinsic Dimension of Texts: from Academic Abstract to Creative Story</i> Pedashenko Vladislav, Laida Kushnareva, Yana Khassan Nibal, Eduard Tulchinskii, Kristian Ku- znetsov, Vladislav Zharchinskii, Yury Maximov and Irina Piontkovskaya	7916
<i>Image Corruption-Inspired Membership Inference Attacks against Large Vision-Language Models</i> Zongyu Wu, Minhua Lin, Zhiwei Zhang, Fali Wang, Xianren Zhang, Xiang Zhang and Suhang Wang	7945
<i>Language Lives in Sparse Dimensions: Toward Interpretable and Efficient Multilingual Control for Large Language Models</i> Chengzhi Zhong, Fei Cheng, Qianying Liu, Yugo Murawaki, Chenhui Chu and Sadao Kurohashi	7958
<i>Engagement Undermines Safety: How Stereotypes and Toxicity Shape Humor in Language Models</i> Atharvan Dogra, Soumya Suvra Ghosal, Ameet Deshpande, Ashwin Kalyan and Dinesh Manocha	7971
<i>Are All Prompt Components Value-Neutral? Understanding the Heterogeneous Adversarial Robustness of Dissected Prompt in LLMs</i> Yujia Zheng, Tianhao Li, Haotian Huang, Tianyu Zeng, Jingyu Lu, Chuangxin Chu, Yuekai Huang, Ziyou Jiang, Qian Xiong, Yuyao Ge and Mingyang Li	7991
<i>A Regex Minimization Benchmark: A PSPACE-Complete Challenge for Language Models</i> Hyundong Jin, Joonghyuk Hahn and Yo-Sub Han	8020
<i>Teaching Small Language Models to Learn Logic through Meta-Learning</i> Leonardo Bertolazzi, Manuel Vargas Guzmán, Raffaella Bernardi, Maciej Malicki and Jakub Szymanik	8049
<i>COMPACT: Building Compliance Paralegals via Clause Graph Reasoning over Contracts</i> Ayush Singh, Dishank Aggarwal, Pranav Bhagat, Ainulla Khan, Sameer Malik and Amar Prakash Azad	8081
<i>Surprisal and Metaphor Novelty Judgments: Moderate Correlations and Divergent Scaling Effects Revealed by Corpus-Based and Synthetic Datasets</i> Omar Momen, Emilie Sitter, Berenike Herrmann and Sina Zarrieß	8113
<i>Repairing Regex Vulnerabilities via Localization-Guided Instructions</i> Sicheol Sung, Joonghyuk Hahn and Yo-Sub Han	8128
<i>Do Psychometric Tests Work for Large Language Models? Evaluation of Tests on Sexism, Racism, and Morality</i> Jana Jung, Marlene Lutz, Indira Sen and Markus Strohmaier	8143

<i>ReFACT: A Benchmark for Scientific Confabulation Detection with Positional Error Annotations</i> Yindong Wang, Martin Preiß, Margarita Bugueño, Jan Vincent Hoffbauer, Abdullatif Ghajar, Tolga Buz and Gerard de Melo	8174
<i>Cosine Similarity as Logits?: A Scalable Knowledge Probe Using Embedding Vectors from Generative Language Models</i> Tomoyuki Jinno, Kazuki Hayashi, Yusuke Sakai, Hidetaka Kamigaito and Taro Watanabe ..	8188
<i>Generating Multi-Aspect Queries for Conversational Search</i> Zahra Abbasiantaeb, Simon Lupart and Mohammad Aliannejadi	8201
<i>Navigating the Infinite Dynamic Web Space: Effective In-Context Exploration via Cognitive Multi- Agent Collaboration</i> Guozhao Mo, Yanjiang Liu, Yafei Shi, Jiawei Chen, Yang Li, Yaojie Lu, Hongyu Lin, Ben He, Le Sun, Bo Zheng and Xianpei Han	8218
<i>TimeMachine-bench: A Benchmark for Evaluating Model Capabilities in Repository-Level Migration Tasks</i> Ryo Fujii, Makoto Morishita, Kazuki Yano and Jun Suzuki	8233
<i>Tandem Training for Language Models</i> Robert West, Ashton Anderson, Ece Kamar and Eric Horvitz	8265
<i>Can MLLMs Find Their Way in a City? Exploring Emergent Navigation from Web-Scale Knowledge</i> Dwip Dalal, Utkarsh Mishra, Narendra Ahuja and Nebojsa Jojic	8279
<i>Wikontic: Constructing Wikidata-Aligned, Ontology-Aware Knowledge Graphs with Large Language Models</i> Alla Chepurova, Aydar Bulatov, Mikhail Burtsev and Yuri Kuratov	8304
<i>CAIRE: Cultural Attribution of Images with Retrieval</i> Arnav Yayavaram, Siddharth Yayavaram, Simran Khanuja, Michael Saxon and Graham Neubig	8320
<i>What Does Infect Mean to Cardio? Investigating the Role of Clinical Specialty Data in Medical LLMs</i> Xinlan Yan, Di Wu, Yibin Lei, Christof Monz and Iacer Calixto	8339
<i>Redefining Retrieval Evaluation in the Era of LLMs</i> Giovanni Trappolini, Florin Cuconasu, Simone Filice, Yoelle Maarek and Fabrizio Silvestri	8359
<i>Debate, Deliberate, Decide (D3): A Cost-Aware Adversarial Framework for Reliable and Interpretable LLM Evaluation</i> Abir Harrasse, Chaithanya Bandi and Hari Bandi	8376
<i>IYKYK: Using language models to decode extremist cryptolects</i> Christine de Kock, Arij Riabi, Zeerak Talat, Michael Sejr Schlichtkrull, Pranava Madhyastha and Eduard Hovy	8393
<i>Stop Taking Tokenizers for Granted: They Are Core Design Decisions in Large Language Models</i> Sawsan Alqahtani, Mir Tafseer Nayeem, Md Tahmid Rahman Laskar, Tasnim Mohiuddin and M Saiful Bari	8410