

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

Short Papers

Bangkok, Thailand
11-16 August 2024

Volume 1 of 2

ISBN: 979-8-3313-0149-1

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>Can Language Models Serve as Text-Based World Simulators?</i> Ruoyao Wang, Graham Todd, Ziang Xiao, Xingdi Yuan, Marc-Alexandre Côté, Peter Clark and Peter Jansen	1
<i>FanOutQA: A Multi-Hop, Multi-Document Question Answering Benchmark for Large Language Models</i> Andrew Zhu, Alyssa Hwang, Liam Dugan and Chris Callison-Burch	18
<i>Revisiting Code Similarity Evaluation with Abstract Syntax Tree Edit Distance</i> Yewei Song, Cedric Lothritz, Xunzhu Tang, Tegawendé F. Bissyandé and Jacques Klein	38
<i>Resisting the Lure of the Skyline: Grounding Practices in Active Learning for Morphological Inflection</i> Saliha Muradoglu, Michael Ginn, Miikka Silfverberg and Mans Hulden	47
<i>Speculative Contrastive Decoding</i> Hongyi Yuan, Keming Lu, Fei Huang, Zheng Yuan and Chang Zhou	56
<i>RDRec: Rationale Distillation for LLM-based Recommendation</i> Xinfeng Wang, Jin Cui, Yoshimi Suzuki and Fumiyo Fukumoto	65
<i>Isotropy, Clusters, and Classifiers</i> Timothee Mickus, Stig-Arne Grönroos and Joseph Attieh	75
<i>Language Models Do Hard Arithmetic Tasks Easily and Hardly Do Easy Arithmetic Tasks</i> Andrew Gambardella, Yusuke Iwasawa and Yutaka Matsuo	85
<i>Simpson’s Paradox and the Accuracy-Fluency Tradeoff in Translation</i> Zheng Wei Lim, Ekaterina Vylomova, Trevor Cohn and Charles Kemp	92
<i>UltraSparseBERT: 99% Conditionally Sparse Language Modelling</i> Peter Belcak and Roger Wattenhofer	104
<i>SceMQA: A Scientific College Entrance Level Multimodal Question Answering Benchmark</i> Zhenwen Liang, Kehan Guo, Gang Liu, Taicheng Guo, Yujun Zhou, Tianyu Yang, Jiajun Jiao, Renjie Pi, Jipeng Zhang and Xiangliang Zhang	109
<i>On the Role of Long-tail Knowledge in Retrieval Augmented Large Language Models</i> Dongyang Li, Junbing Yan, Taolin Zhang, Chengyu Wang, Xiaofeng He, Longtao Huang, Hui Xue’ and Jun Huang	120
<i>IEPile: Unearthing Large Scale Schema-Conditioned Information Extraction Corpus</i> Honghao Gui, Lin Yuan, Hongbin Ye, Ningyu Zhang, Mengshu Sun, Lei Liang and Huajun Chen	127
<i>Bi-Directional Multi-Granularity Generation Framework for Knowledge Graph-to-Text with Large Language Model</i> Haowei Du, Chen Li, Dinghao Zhang and Dongyan Zhao	147
<i>Code-Switching Can be Better Aligners: Advancing Cross-Lingual SLU through Representation-Level and Prediction-Level Alignment</i> Zhihong Zhu, Xuxin Cheng, Zhanpeng Chen, Xianwei Zhuang, Zhiqi Huang and Yuexian Zou	153

<i>AFLoRA: Adaptive Freezing of Low Rank Adaptation in Parameter Efficient Fine-Tuning of Large Models</i>	
Zeyu Liu, Souvik Kundu, Anni Li, Junrui Wan, Lianghao Jiang and Peter Anthony Beerel . . .	161
<i>DDPrompt: Differential Diversity Prompting in Large Language Models</i>	
Lin Mu, Wenhao Zhang, Yiwen Zhang and Peiquan Jin	168
<i>Monotonic Representation of Numeric Attributes in Language Models</i>	
Benjamin Heinzerling and Kentaro Inui	175
<i>Two Issues with Chinese Spelling Correction and A Refinement Solution</i>	
Changxuan Sun, Linlin She and Xuesong Lu	196
<i>DynaSemble: Dynamic Ensembling of Textual and Structure-Based Models for Knowledge Graph Completion</i>	
Ananjan Nandi, Navdeep Kaur, Parag Singla and Mausam	205
<i>Fine-Tuning Pre-Trained Language Models with Gaze Supervision</i>	
Shuwen Deng, Paul Prasse, David Robert Reich, Tobias Scheffer and Lena Ann Jäger	217
<i>Growing Trees on Sounds: Assessing Strategies for End-to-End Dependency Parsing of Speech</i>	
Adrien Pupier, Maximin Coavoux, Jérôme Goulian and Benjamin Lecouteux	225
<i>Sketch-Guided Constrained Decoding for Boosting Blackbox Large Language Models without Logit Access</i>	
Saibo Geng, Berkay Döner, Chris Wendler, Martin Josifoski and Robert West	234
<i>On the Semantic Latent Space of Diffusion-Based Text-To-Speech Models</i>	
Miri Varshavsky-Hassid, Roy Hirsch, Regev Cohen, Tomer Golany, Daniel Freedman and Ehud Rivlin	246
<i>Learnable Privacy Neurons Localization in Language Models</i>	
Ruizhe Chen, Tianxiang Hu, Yang Feng and Zuozhu Liu	256
<i>Is the Pope Catholic? Yes, the Pope is Catholic. Generative Evaluation of Non-Literal Intent Resolution in LLMs</i>	
Akhila Yerukola, Saujas Vaduguru, Daniel Fried and Maarten Sap	265
<i>Generating Harder Cross-document Event Coreference Resolution Datasets using Metaphoric Paraphrasing</i>	
Shafiuddin Rehan Ahmed, Zhiyong Wang, George Arthur Baker, Kevin Stowe and James H. Martin	276
<i>Soft Self-Consistency Improves Language Models Agents</i>	
Han Wang, Archiki Prasad, Elias Stengel-Eskin and Mohit Bansal	287
<i>RecGPT: Generative Pre-training for Text-based Recommendation</i>	
Hoang Ngo and Dat Quoc Nguyen	302
<i>MTP: A Dataset for Multi-Modal Turning Points in Casual Conversations</i>	
Gia-Bao Dinh Ho, Chang Wei Tan, Zahra Zamanzadeh Darban, Mahsa Salehi, Reza Haf and Wray Buntine	314
<i>What Does Parameter-free Probing Really Uncover?</i>	
Tommi Buder-Gröndahl	327
<i>ATLAS: Improving Lay Summarisation with Attribute-based Control</i>	
Zhihao Zhang, Tomas Goldsack, Carolina Scarton and Chenghua Lin	337

<i>EmbSpatial-Bench: Benchmarking Spatial Understanding for Embodied Tasks with Large Vision-Language Models</i>	
Mengfei Du, Binhao Wu, Zejun Li, Xuanjing Huang and Zhongyu Wei	346
<i>Understanding the Effects of Noise in Text-to-SQL: An Examination of the BIRD-Bench Benchmark</i>	
Niklas Wretblad, Fredrik Gordh Riseby, Rahul Biswas, Amin Ahmadi and Oskar Holmström	356
<i>Dwell in the Beginning: How Language Models Embed Long Documents for Dense Retrieval</i>	
João Coelho, Bruno Martins, Joao Magalhaes, Jamie Callan and Chenyan Xiong	370
<i>That’s Optional: A Contemporary Exploration of thatOmission in English Subordinate Clauses</i>	
Ella Rabinovich	378
<i>Do Large Language Models Discriminate in Hiring Decisions on the Basis of Race, Ethnicity, and Gender?</i>	
Haozhe An, Christabel Acquaye, Colin Wang, Zongxia Li and Rachel Rudinger	386
<i>Explainability and Hate Speech: Structured Explanations Make Social Media Moderators Faster</i>	
Agostina Calabrese, Leonardo Neves, Neil Shah, Maarten W. Bos, Björn Ross, Mirella Lapata and Francesco Barbieri	398
<i>Born Differently Makes a Difference: Counterfactual Study of Bias in Biography Generation from a Data-to-Text Perspective</i>	
Biaoyan Fang, Ritvik Dinesh, Xiang Dai and Sarvnaz Karimi	409
<i>Sign Language Translation with Sentence Embedding Supervision</i>	
Hamidullah Yasser, Josef Van Genabith and Cristina España-Bonet	425
<i>STREAM: Simplified Topic Retrieval, Exploration, and Analysis Module</i>	
Anton Frederik Thielmann, Arik Reuter, Christoph Weisser, Gillian Kant, Manish Kumar and Benjamin Säfken	435
<i>DocFinQA: A Long-Context Financial Reasoning Dataset</i>	
Varshini Reddy, Rik Koncel-Kedziorski, Viet Dac Lai, Michael Krumdick, Charles Lovering and Chris Tanner	445
<i>MaskLID: Code-Switching Language Identification through Iterative Masking</i>	
Amir Hossein Kargaran, François Yvon and Hinrich Schuetze	459
<i>An Empirical Analysis on Large Language Models in Debate Evaluation</i>	
Xinyi Liu, Pinxin Liu and Hangfeng He	470
<i>Fine-Tuned Machine Translation Metrics Struggle in Unseen Domains</i>	
Vilém Zouhar, Shuoyang Ding, Anna Currey, Tatyana Badeka, Jenyuan Wang and Brian Thompson	488
<i>IndicIRSuite: Multilingual Dataset and Neural Information Models for Indian Languages</i>	
Saiful Haq, Ashutosh Sharma, Omar Khattab, Niyati Chhaya and Pushpak Bhattacharyya . . .	501
<i>AGR: Reinforced Causal Agent-Guided Self-explaining Rationalization</i>	
Yunxiao Zhao, Zhiqiang Wang, Xiaoli Li, Jiye Liang and Ru Li	510
<i>Shoulders of Giants: A Look at the Degree and Utility of Openness in NLP Research</i>	
Surangika Ranathunga, Nisansa De Silva, Dilith Jayakody and Aloka Fernando	519
<i>The Probabilities Also Matter: A More Faithful Metric for Faithfulness of Free-Text Explanations in Large Language Models</i>	
Noah Yamamoto Siegel, Oana-Maria Camburu, Nicolas Heess and Maria Perez-Ortiz	530

<i>Naming, Describing, and Quantifying Visual Objects in Humans and LLMs</i> Alberto Testoni, Juell Sprott and Sandro Pezzelle	547
<i>Are LLMs classical or nonmonotonic reasoners? Lessons from generics</i> Alina Leiding, Robert Van Rooij and Ekaterina Shutova	558
<i>ConstitutionalExperts: Training a Mixture of Principle-based Prompts</i> Savvas Petridis, Ben Wedin, Ann Yuan, James Wexler and Nithum Thain	574
<i>Time Sensitive Knowledge Editing through Efficient Finetuning</i> Xiou Ge, Ali Mousavi, Edouard Grave, Armand Joulin, Kun Qian, Benjamin Han, Mostafa Arefiyan and Yunyao Li	583
<i>PRewrite: Prompt Rewriting with Reinforcement Learning</i> Weize Kong, Spurthi Amba Hombaiah, Mingyang Zhang, Qiaozhu Mei and Michael Bendersky	594
<i>Paraphrasing in Affirmative Terms Improves Negation Understanding</i> MohammadHossein Rezaei and Eduardo Blanco	602
<i>Exploring Conditional Variational Mechanism to Pinyin Input Method for Addressing One-to-Many Mappings in Low-Resource Scenarios</i> Bin Sun, Jianfeng Li, Hao Zhou, Fandong Meng, Kan Li and Jie Zhou	616
<i>Consistency Training by Synthetic Question Generation for Conversational Question Answering</i> Hamed Hematian Hemati and Hamid Beigy	630
<i>How Good is Zero-Shot MT Evaluation for Low Resource Indian Languages?</i> Anushka Singh, Ananya B. Sai, Raj Dabre, Ratish Puduppully, Anoop Kunchukuttan and Mitesh M Khapra	640
<i>Zero-Shot Cross-Lingual Reranking with Large Language Models for Low-Resource Languages</i> Mofetoluwa Adeyemi, Akintunde Oladipo, Ronak Pradeep and Jimmy Lin	650
<i>Cross-Modal Projection in Multimodal LLMs Doesn't Really Project Visual Attributes to Textual Space</i> Gaurav Verma, Minje Choi, Kartik Sharma, Jamelle Watson-Daniels, Sejoon Oh and Srijan Kumar	657
<i>Guidance-Based Prompt Data Augmentation in Specialized Domains for Named Entity Recognition</i> Hyeonseok Kang, Hyein Seo, Jeesu Jung, Sangkeun Jung, Du-Seong Chang and Riwoo Chung	665
<i>Aligning Large Language Models via Fine-grained Supervision</i> Dehong Xu, Liang Qiu, Minseok Kim, Faisal Ladhak and Jaeyoung Do	673
<i>Annotating FrameNet via Structure-Conditioned Language Generation</i> Xinyue Cui and Swabha Swayamdipta	681
<i>DUAL-REFLECT: Enhancing Large Language Models for Reflective Translation through Dual Learning Feedback Mechanisms</i> Andong Chen, Lianzhang Lou, Kehai Chen, Xuefeng Bai, Yang Xiang, Muyun Yang, Tiejun Zhao and Min Zhang	693
<i>Towards Artwork Explanation in Large-scale Vision Language Models</i> Kazuki Hayashi, Yusuke Sakai, Hidetaka Kamigaito, Katsuhiko Hayashi and Taro Watanabe	705
<i>On the Hallucination in Simultaneous Machine Translation</i> Meizhi Zhong, Kehai Chen, Zhengshan Xue, Lemao Liu, Mingming Yang and Min Zhang ..	730

<i>Self-Augmented In-Context Learning for Unsupervised Word Translation</i> Yaoyiran Li, Anna Korhonen and Ivan Vulić	743
<i>RAM-EHR: Retrieval Augmentation Meets Clinical Predictions on Electronic Health Records</i> Ran Xu, Wenqi Shi, Yue Yu, Yuchen Zhuang, Bowen Jin, May Dongmei Wang, Joyce C. Ho and Carl Yang	754
<i>Estimating the Level of Dialectness Predicts Inter-annotator Agreement in Multi-dialect Arabic Data-sets</i> Amr Keleg, Walid Magdy and Sharon Goldwater	766
<i>Estimating the Level of Dialectness Predicts Inter-annotator Agreement in Multi-dialect Arabic Data-sets</i> Amr Keleg, Walid Magdy and Sharon Goldwater	778
<i>Linear-time Minimum Bayes Risk Decoding with Reference Aggregation</i> Jannis Vamvas and Rico Sennrich	790
<i>Cleaner Pretraining Corpus Curation with Neural Web Scraping</i> Zhipeng Xu, Zhenghao Liu, Yukun Yan, Zhiyuan Liu, Ge Yu and Chenyan Xiong	802
<i>Greed is All You Need: An Evaluation of Tokenizer Inference Methods</i> Omri Uzan, Craig W. Schmidt, Chris Tanner and Yuval Pinter	813
<i>What Do Dialect Speakers Want? A Survey of Attitudes Towards Language Technology for German Dialects</i> Verena Blaschke, Christoph Purschke, Hinrich Schuetze and Barbara Plank	823
<i>SeeGULL Multilingual: a Dataset of Geo-Culturally Situated Stereotypes</i> Mukul Bhutani, Kevin Robinson, Vinodkumar Prabhakaran, Shachi Dave and Sunipa Dev ...	842
<i>Getting Serious about Humor: Crafting Humor Datasets with Unfunny Large Language Models</i> Zachary Horvitz, Jingru Chen, Rahul Aditya, Harshvardhan Srivastava, Robert West, Zhou Yu and Kathleen McKeown	855
<i>Don't Buy it! Reassessing the Ad Understanding Abilities of Contrastive Multimodal Models</i> Anna Bavaresco, Alberto Testoni and Raquel Fernández	870