

Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)

Industry Track

Miami, Florida, USA
12-16 November 2024

Volume 1 of 2

ISBN: 979-8-3313-0839-1

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. (2025)

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>Optimizing Entity Resolution in Voice Interfaces: An ASR-Aware Entity Reference Expansion Approach</i> Jiangning Chen, Ziyun Zhang and Qianli Hu	1
<i>Two-tiered Encoder-based Hallucination Detection for Retrieval-Augmented Generation in the Wild</i> Ilana Zimmerman, Jadin Tredup, Ethan Selfridge and Joseph Bradley	8
<i>The Program Testing Ability of Large Language Models for Code</i> Weimin Xiong, Yiwen Guo and Hao Chen	23
<i>Salient Information Prompting to Steer Content in Prompt-based Abstractive Summarization</i> Lei Xu, Mohammed Asad Karim, Saket Dingliwal and Aparna Elangovan	35
<i>Predicting Entity Saliency in Extremely Short Documents</i> Benjamin Bullough, Harrison Lundberg, Chen Hu and Weihang Xiao	50
<i>Don't Shoot The Breeze: Topic Continuity Model Using Nonlinear Naive Bayes With Attention</i> Shu-Ting Pi, Pradeep Bagavan, Yejia Li, Disha Disha and Qun Liu	65
<i>Retrieval Augmented Spelling Correction for E-Commerce Applications</i> Xuan Guo, Rohit Patki, Dante Everaert and Christopher Potts	73
<i>Scaling Parameter-Constrained Language Models with Quality Data</i> Ernie Chang, Matteo Paltenghi, Yang Li, Pin-Jie Lin, Changsheng Zhao, Patrick Huber, Zechun Liu, Rastislav Rabatin, Yangyang Shi and Vikas Chandra	80
<i>INDUS: Effective and Efficient Language Models for Scientific Applications</i> Bishwaranjan Bhattacharjee, Aashka Trivedi, Masayasu Muraoka, Muthukumaran Ramasubramanian, Takuma Udagawa, Iksha Gurung, Nishan Pantha, Rong Zhang, Bharath Dandala, Rahul Ramachandran, Manil Maskey, Kaylin Bugbee, Michael M. Little, Elizabeth Fancher, Irina Gerasimov, Armin Mehrabian, Lauren Sanders, Sylvain V. Costes, Sergi Blanco-Cuaresma, Kelly Lockhart, Thomas Allen, Felix Grezes, Megan Ansdell, Alberto Accomazzi, Yousef El-Kurdi, Davis Wertheimer, Birgit Pfitzmann, Cesar Berrospi Ramis, Michele Dolfi, Rafael Teixeira De Lima, Panagiotis Vagenas, S. Karthik Mukkavilli, Peter W. J. Staar, Sanaz Vahidinia, Ryan McGranaghan and Tsengdar J. Lee	98
<i>DL-QAT: Weight-Decomposed Low-Rank Quantization-Aware Training for Large Language Models</i> Wenjing Ke, Zhe Li, Dong Li, Lu Tian and Emad Barsoum	113
<i>Hybrid-RACA: Hybrid Retrieval-Augmented Composition Assistance for Real-time Text Prediction</i> Menglin Xia, Xuchao Zhang, Camille Couturier, Guoqing Zheng, Saravan Rajmohan and Victor Rühle	120
<i>LLMC: Benchmarking Large Language Model Quantization with a Versatile Compression Toolkit</i> Ruihao Gong, Yang Yong, Shiqiao Gu, Yushi Huang, Chengtao Lv, Yunchen Zhang, Dacheng Tao and Xianglong Liu	132
<i>PDFTriage: Question Answering over Long, Structured Documents</i> Jon Saad-Falcon, Joe Barrow, Alexa Siu, Ani Nenkova, Seunghyun Yoon, Ryan A. Rossi and Franck Dernoncourt	153
<i>Fairness-Aware Online Positive-Unlabeled Learning</i> Hoin Jung and Xiaoqian Wang	170

<i>SAAS: Solving Ability Amplification Strategy for Enhanced Mathematical Reasoning in Large Language Models</i>	
Hyeonwoo Kim, Gyoungjin Gim, Yungi Kim, Jihoo Kim, Byungju Kim, Wonseok Lee and Chanjun Park	186
<i>Debiasing Text Safety Classifiers through a Fairness-Aware Ensemble</i>	
Olivia Sturman, Aparna R Joshi, Bhaktipriya Radharapu, Piyush Kumar and Renee Shelby . .	199
<i>Centrality-aware Product Retrieval and Ranking</i>	
Hadeel Saadany, Swapnil Bhosale, Samarth Agrawal, Diptesh Kanojia, Constantin Orasan and Zhe Wu	215
<i>Fusion-Eval: Integrating Assistant Evaluators with LLMs</i>	
Lei Shu, Nevan Wichers, Liangchen Luo, Yun Zhu, Yinxiao Liu, Jindong Chen and Lei Meng	225
<i>Investigating the Personality Consistency in Quantized Role-Playing Dialogue Agents</i>	
Yixiao Wang, Homa Fashandi and Kevin Ferreira	239
<i>Robust ASR Error Correction with Conservative Data Filtering</i>	
Takuma Udagawa, Masayuki Suzuki, Masayasu Muraoka and Gakuto Kurata	256
<i>Code Representation Pre-training with Complements from Program Executions</i>	
Jiabo Huang, Jianyu Zhao, Yuyang Rong, Yiwen Guo, Yifeng He and Hao Chen	267
<i>ScaleLLM: A Resource-Frugal LLM Serving Framework by Optimizing End-to-End Efficiency</i>	
Yuhang Yao, Han Jin, Alay Dilipbhai Shah, Shanshan Han, Zijian Hu, Dimitris Stripelis, Yide Ran, Zhaozhuo Xu, Salman Avestimehr and Chaoyang He	279
<i>Context Matters: Pushing the Boundaries of Open-Ended Answer Generation with Graph-Structured Knowledge Context</i>	
Somnath Banerjee, Amrui Sahoo, Sayan Layek, Avik Dutta, Rima Hazra and Animesh Mukherjee	290
<i>SHIELD: LLM-Driven Schema Induction for Predictive Analytics in EV Battery Supply Chain Disruptions</i>	
Zhi-Qi Cheng, Yifei Dong, Aike Shi, Wei Liu, Yuzhi Hu, Jason O'Connor, Alexander G Hauptmann and Kate Whitefoot	303
<i>Divide-Conquer-Reasoning for Consistency Evaluation and Automatic Improvement of Large Language Models</i>	
Wendi Cui, Zhuohang Li, Damien Lopez, Kamalika Das, Bradley A. Malin, Sricharan Kumar and Jiaxin Zhang	334
<i>Greenback Bears and Fiscal Hawks: Finance is a Jungle and Text Embeddings Must Adapt</i>	
Peter Anderson, Mano Vikash Janardhanan, Jason He, Wei Cheng and Charlie Flanagan . . .	362
<i>TPTU-v2: Boosting Task Planning and Tool Usage of Large Language Model-based Agents in Real-world Industry Systems</i>	
Yilun Kong, Jingqing Ruan, YiHong Chen, Bin Zhang, Tianpeng Bao, Shi Shiwei, du Guo Qing, Xiaoru Hu, Hangyu Mao, Ziyue Li, Xingyu Zeng, Rui Zhao and Xueqian Wang	371
<i>Detecting Ambiguous Utterances in an Intelligent Assistant</i>	
Satoshi Akasaki and Manabu Sassano	386
<i>GeoIndia: A Seq2Seq Geocoding Approach for Indian Addresses</i>	
Bhavuk Singh, Anshu Aditya, Lokesh Todwal, Shubham Jain and Debashis Mukherjee . . .	395

<i>Moleco: Molecular Contrastive Learning with Chemical Language Models for Molecular Property Prediction</i>	
Jun-Hyung Park, Hyuntae Park, Yeachan Kim, Woosang Lim and SangKeun Lee	408
<i>SEED: Semantic Knowledge Transfer for Language Model Adaptation to Materials Science</i>	
Yeachan Kim, Jun-Hyung Park, SungHo Kim, Juhyeong Park, Sangyun Kim and SangKeun Lee	421
<i>News Risk Alerting System (NRAS): A Data-Driven LLM Approach to Proactive Credit Risk Monitoring</i>	
Adil Nygaard, Ashish Upadhyay, Lauren Hinkle, Xenia Skotti, Joe Halliwell, Ian C Brown and Glen Noronha	429
<i>FastAdaSP: Multitask-Adapted Efficient Inference for Large Speech Language Model</i>	
Yichen Lu, Jiaqi Song, Chao-Han Huck Yang and Shinji Watanabe	440
<i>TensorOpera Router: A Multi-Model Router for Efficient LLM Inference</i>	
Dimitris Stripelis, Zhaozhuo Xu, Zijian Hu, Alay Dilipbhai Shah, Han Jin, Yuhang Yao, Jipeng Zhang, Tong Zhang, Salman Avestimehr and Chaoyang He	452
<i>Prompt-Tuned Multi-Task Taxonomic Transformer (PTMTTaxoFormer)</i>	
Rajashekar Vasantha, Nhan Nguyen and Yue Zhang	463
<i>Arcee's MergeKit: A Toolkit for Merging Large Language Models</i>	
Charles Goddard, Shamane Siriwardhana, Malikeh Ehghaghi, Luke Meyers, Vladimir Karpukhin, Brian Benedict, Mark McQuade and Jacob Solawetz	477
<i>Personal Large Language Model Agents: A Case Study on Tailored Travel Planning</i>	
Harmanpreet Singh, Nikhil Verma, Yixiao Wang, Manasa Bharadwaj, Homa Fashandi, Kevin Ferreira and Chul Lee	486
<i>FanLoRA: Fantastic LoRAs and Where to Find Them in Large Language Model Fine-tuning</i>	
Aaron Xuxiang Tian, Yi Zhao, Congrui Yin, Wei Zhu, Xing Tian and Yi Ge	515
<i>ReportGPT: Human-in-the-loop Verifiable Table-to-Text Generation</i>	
Lucas Cecchi and Petr Babkin	529
<i>BPID: A Benchmark for Personal Identity Deduplication</i>	
Runhui Wang, Yefan Tao, Adit Krishnan, Luyang Kong, Xuanqing Liu, Yuqian Deng, Yunzhao Yang, Henrik Johnson, Andrew Borthwick, Shobhit Gupta, Aditi Sinha Gundlapalli and Davor Golac	538
<i>MERLIN: Multimodal Embedding Refinement via LLM-based Iterative Navigation for Text-Video Retrieval-Rerank Pipeline</i>	
Donghoon Han, Eunhwan Park, Gisang Lee, Adam Lee and Nojun Kwak	547
<i>Identifying High Consideration E-Commerce Search Queries</i>	
Zhiyu Chen, Jason Ingyu Choi, Besnik Fetahu and Shervin Malmasi	563
<i>Sample Design Engineering: An Empirical Study on Designing Better Fine-Tuning Samples for Information Extraction with LLMs</i>	
Biyang Guo, He Wang, Wenying Xiao, Hong Chen, ZhuXin Lee, Songqiao Han and Hailiang Huang	573
<i>Refining App Reviews: Dataset, Methodology, and Evaluation</i>	
Amrita Singh, Chirag Jain, Mohit Chaudhary and Preethu Rose Anish	595

<i>TelBench: A Benchmark for Evaluating Telco-Specific Large Language Models</i>	
Sunwoo Lee, Dhammiko Arya, Seung-Mo Cho, Gyoung-eun Han, Seokyoung Hong, Wonbeom Jang, Seojin Lee, Sohee Park, Sereimony Sek, Injee Song, Sungbin Yoon and Eric Davis	609
<i>RRADistill: Distilling LLMs' Passage Ranking Ability for Long-Tail Queries Document Re-Ranking on a Search Engine</i>	
Nayoung Choi, Youngjune Lee, Gyu-Hwung Cho, Haeyu Jeong, Jungmin Kong, Saehun Kim, Keunchan Park, Sarah Cho, Inchang Jeong, Gyohee Nam, Sunghoon Han, Wonil Yang and Jaeho Choi	627
<i>KorSmishing Explainer: A Korean-centric LLM-based Framework for Smishing Detection and Explanation Generation</i>	
Yunseung Lee and Daehee Han	642
<i>Time Matters: An End-to-End Solution for Temporal Claim Verification</i>	
Anab Maulana Barik, Wynne Hsu and Mong-Li Lee	657
<i>MILD Bot: Multidisciplinary Childhood Cancer Survivor Question-Answering Bot</i>	
Mirae Kim, Kyubum Hwang, Hayoung Oh, Min Ah Kim, Chaerim Park, Yehwi Park and Chungyeon Lee	665
<i>Breaking the Hourglass Phenomenon of Residual Quantization: Enhancing the Upper Bound of Generative Retrieval</i>	
Zhirui Kuai, Zuxu Chen, Huimu Wang, Mingming Li, Dadong Miao, Wang Binbin, Xusong Chen, Li Kuang, Yuxing Han, Jiaxing Wang, Guoyu Tang, Lin Liu, Songlin Wang and Jingwei Zhuo	677
<i>Improving Few-Shot Cross-Domain Named Entity Recognition by Instruction Tuning a Word-Embedding based Retrieval Augmented Large Language Model</i>	
Subhadip Nandi and Neeraj Agrawal	686
<i>IPL: Leveraging Multimodal Large Language Models for Intelligent Product Listing</i>	
Kang Chen, Qing Heng Zhang, Chengbao Lian, Yixin Ji, Xuwei Liu, Shuguang Han, Guoqiang Wu, Fei Huang and Jufeng Chen	697
<i>QDyLoRA: Quantized Dynamic Low-Rank Adaptation for Efficient Large Language Model Tuning</i>	
Hossein Rajabzadeh, Mojtaba Valipour, Tianshu Zhu, Marzieh S. Tahaei, Hyock Ju Kwon, Ali Ghodsi, Boxing Chen and Mehdi Rezagholizadeh	712
<i>Improving Hierarchical Text Clustering with LLM-guided Multi-view Cluster Representation</i>	
Anup Pattnaik, Cijo George, Rishabh Kumar Tripathi, Sasanka Vutla and Jithendra Vepa	719
<i>PARA: Parameter-Efficient Fine-tuning with Prompt-Aware Representation Adjustment</i>	
Zequan Liu, Yi Zhao, Ming Tan, Wei Zhu and Aaron Xuxiang Tian	728
<i>RAG4ITOps: A Supervised Fine-Tunable and Comprehensive RAG Framework for IT Operations and Maintenance</i>	
Tianyang Zhang, Zhuoxuan Jiang, Shengguang Bai, Tianrui Zhang, Lin Lin, Yang Liu and Jiawei Ren	738
<i>ULMR: Unlearning Large Language Models via Negative Response and Model Parameter Average</i>	
Shaojie Shi, Xiaoyu Tan, Xihe Qiu, Chao Qu, Kexin Nie, Yuan Cheng, Wei Chu, Xu Yinghui and Yuan Qi	755
<i>Pretraining and Finetuning Language Models on Geospatial Networks for Accurate Address Matching</i>	
Saket Maheshwary, Arpan Paul and Saurabh Sohoney	763

<i>SMARTCAL: An Approach to Self-Aware Tool-Use Evaluation and Calibration</i> Yuanhao Shen, Xiaodan Zhu and Lei Chen	774
<i>Probing the Depths of Language Models' Contact-Center Knowledge for Quality Assurance</i> Digvijay Anil Ingle, Aashraya Sachdeva, Surya Prakash Sahu, Mayank Sati, Cijo George and Jithendra Vepa	790
<i>Intelligent Predictive Maintenance RAG framework for Power Plants: Enhancing QA with StyleDFS and Domain Specific Instruction Tuning</i> Seongtae Hong, Joong Min Shin, Jaehyung Seo, Taemin Lee, Jeongbae Park, Cho Man Young, Byeongho Choi and Heuseok Lim	805
<i>Structured Object Language Modeling (SO-LM): Native Structured Objects Generation Conforming to Complex Schemas with Self-Supervised Denoising</i> Amir Tavanaei, Kee Kiat Koo, Hayreddin Ceker, Shaobai Jiang, Qi Li, Julien Han and Karim Bouyarmane	821
<i>Assisting Breastfeeding and Maternity Experts in Responding to User Queries with an AI-in-the-loop Approach</i> Nadjet Bouayad-Agha, Ignasi Gomez-Sebastia, Alba Padro, Enric Pallares Roura, David Pelayo Castelló and Rocío Tovar	829
<i>A Hassle-free Algorithm for Strong Differential Privacy in Federated Learning Systems</i> Hugh Brendan McMahan, Zheng Xu and Yanxiang Zhang	842
<i>ProConSuL: Project Context for Code Summarization with LLMs</i> Vadim Lomshakov, Andrey Podivilov, Sergey Savin, Oleg Baryshnikov, Alena Lisevych and Ser- gey Nikolenko	866
<i>Retrieval Augmented Generation or Long-Context LLMs? A Comprehensive Study and Hybrid Approa- ch</i> Zhuowan Li, Cheng Li, Mingyang Zhang, Qiaozhu Mei and Michael Bendersky	881
<i>MARS: Multilingual Aspect-centric Review Summarisation</i> Sandeep Sricharan Mukku, Abinеш Kanagarajan, Chetan Aggarwal and Promod Yenigalla ..	894
<i>A new approach for fine-tuning sentence transformers for intent classification and out-of-scope detec- tion tasks</i> Tianyi Zhang, Atta Norouzian, Aanchan Mohan and Frederick Ducatelle	910
<i>Tell me what I need to know: Exploring LLM-based (Personalized) Abstractive Multi-Source Meeting Summarization</i> Frederic Kirstein, Terry Ruas, Robert Kratel and Bela Gipp	920
<i>Detecting LLM-Assisted Cheating on Open-Ended Writing Tasks on Language Proficiency Tests</i> Chenhao Niu, Kevin P. Yancey, Ruidong Liu, Mirza Basim Baig, André Kenji Horie and James Sharpnack	940
<i>Can Machine Unlearning Reduce Social Bias in Language Models?</i> Omkar Dige, Diljot Arneja, Tsz Fung Yau, Qixuan Zhang, Mohammad Bolandraftar, Xiaodan Zhu and Faiza Khan Khattak	954
<i>Don't be my Doctor! Recognizing Healthcare Advice in Large Language Models</i> Kellen Tan Cheng, Anna Lisa Gentile, Pengyuan Li, Chad DeLuca and Guang-Jie Ren	970

<i>Building an Efficient Multilingual Non-Profit IR System for the Islamic Domain Leveraging Multiprocessing Design in Rust</i>	
Vera Pavlova and Mohammed Makhlouf	981
<i>Adapting LLMs for Structured Natural Language API Integration</i>	
Robin Chan, Katsiaryna Mirylenka, Thomas Gschwind, Christoph Miksovic, Paolo Scotton, Enrico Toniato and Abdel Labbi	991
<i>OMG-QA: Building Open-Domain Multi-Modal Generative Question Answering Systems</i>	
Linyong Nan, Weining Fang, Aylin Rasteh, Pouya Lahabi, Weijin Zou, Yilun Zhao and Arman Cohan	1001
<i>Survival of the Safest: Towards Secure Prompt Optimization through Interleaved Multi-Objective Evolution</i>	
Ankita Sinha, Wendi Cui, Kamalika Das and Jiaxin Zhang	1016
<i>Fine-Tuning Large Language Models for Stock Return Prediction Using Newsflow</i>	
Tian Guo and Emmanuel Hauptmann	1028
<i>AmazonQAC: A Large-Scale, Naturalistic Query Autocomplete Dataset</i>	
Dante Everaert, Rohit Patki, Tianqi Zheng and Christopher Potts	1046
<i>Language, OCR, Form Independent (LOFI) pipeline for Industrial Document Information Extraction</i>	
Chang Oh Yoon, Wonbeen Lee, Seokhwan Jang, Kyuwon Choi, Minsung Jung and Daewoo Choi	1056
<i>The State of the Art of Large Language Models on Chartered Financial Analyst Exams</i>	
Mahmoud Mahfouz, Ethan Callanan, Mathieu Sibue, Antony Papadimitriou, Zhiqiang Ma, Xiaomo Liu and Xiaodan Zhu	1068
<i>Value Alignment from Unstructured Text</i>	
Inkit Padhi, Karthikeyan Natesan Ramamurthy, Prasanna Sattigeri, Manish Nagireddy, Pierre Dognin and Kush R. Varshney	1083
<i>LARA: Linguistic-Adaptive Retrieval-Augmentation for Multi-Turn Intent Classification</i>	
Junhua Liu, Tan Yong Keat, Bin Fu and Kwan Hui Lim	1096
<i>Generating Vehicular Icon Descriptions and Indications Using Large Vision-Language Models</i>	
James Fletcher, Nicholas Dehnen, Seyed Nima Tayarani Bathaie, Aijun An, Heidar Davoudi, Ron DiCarlantonio and Gary Farmaner	1107
<i>Athena: Safe Autonomous Agents with Verbal Contrastive Learning</i>	
Tanmana Sadhu, Ali Pesaranghader, Yanan Chen and Dong Hoon Yi	1121
<i>Granite-Function Calling Model: Introducing Function Calling Abilities via Multi-task Learning of Granular Tasks</i>	
Ibrahim Abdelaziz, Kinjal Basu, Mayank Agarwal, Sadhana Kumaravel, Matthew Stallone, Rameswar Panda, Yara Rizk, G P Shrivatsa Bhargav, Maxwell Crouse, Chulaka Gunasekara, Shajith Iqbal, Sachindra Joshi, Hima Karanam, Vineet Kumar, Asim Munawar, Sumit Neelam, Dinesh Raghu, Udit Sharma, Adriana Meza Soria, Dheeraj Sreedhar, Praveen Venkateswaran, Merve Unuvar, David Daniel Cox, Salim Roukos, Luis A. Lastras and Pavan Kapanipathi	1131
<i>Query-OPT: Optimizing Inference of Large Language Models via Multi-Query Instructions in Meeting Summarization</i>	
Md Tahmid Rahman Laskar, Elena Khasanova, Xue-Yong Fu, Cheng Chen and Shashi Bhushan Tn	1140

<i>DiAL : Diversity Aware Listwise Ranking for Query Auto-Complete</i> Sonali Singh, Sachin Sudhakar Farfade and Prakash Mandayam Comar	1152
<i>Systematic Evaluation of Long-Context LLMs on Financial Concepts</i> Lavanya Gupta, Saket Sharma and Yiyun Zhao	1163
<i>ConvKGYarn: Spinning Configurable and Scalable Conversational Knowledge Graph QA Datasets with Large Language Models</i> Ronak Pradeep, Daniel Lee, Ali Mousavi, Jeffrey Pound, Yisi Sang, Jimmy Lin, Ihab Ilyas, Saloni Potdar, Mostafa Arefiyan and Yunyao Li	1176
<i>Knowledge-augmented Financial Market Analysis and Report Generation</i> Yuemin Chen, Feifan Wu, Jingwei Wang, Hao Qian, Ziqi Liu, Zhiqiang Zhang, Jun Zhou and Meng Wang	1207
<i>Let Me Speak Freely? A Study On The Impact Of Format Restrictions On Large Language Model Performance.</i> Zhi Rui Tam, Cheng-Kuang Wu, Yi-Lin Tsai, Chieh-Yen Lin, Hung-yi Lee and Yun-Nung Chen	1218
<i>ASTRA: Automatic Schema Matching using Machine Translation</i> Tarang Chugh and Deepak Zambre	1237
<i>Neural Search Space in Gboard Decoder</i> Yanxiang Zhang, Yuanbo Zhang, Haicheng Sun, Yun Wang, Gary Sivek and Shumin Zhai .	1245
<i>Prompt Leakage effect and mitigation strategies for multi-turn LLM Applications</i> Divyansh Agarwal, Alexander Fabbri, Ben Risher, Philippe Laban, Shafiq Joty and Chien-Sheng Wu	1255
<i>Sequential LLM Framework for Fashion Recommendation</i> Han Liu, Xianfeng Tang, Tianlang Chen, Jiapeng Liu, Indu Indu, Henry Peng Zou, Peng Dai, Roberto Fernandez Galan, Michael D Porter, Dongmei Jia, Ning Zhang and Lian Xiong	1276
<i>Visual Editing with LLM-based Tool Chaining: An Efficient Distillation Approach for Real-Time Applications</i> Oren Sultan, Alexander Khasin, Guy Shiran, Asnat Greenstein-Messica and Dafna Shahaf .	1286
<i>Provenance: A Light-weight Fact-checker for Retrieval Augmented LLM Generation Output</i> Hithesh Sankararaman, Mohammed Nasheed Yasin, Tanner Sorensen, Alessandro Di Bari and Andreas Stolcke	1305
<i>AnyMAL: An Efficient and Scalable Any-Modality Augmented Language Model</i> Seungwhan Moon, Andrea Madotto, Zhaojiang Lin, Tushar Nagarajan, Matt Smith, Shashank Jain, Chun-Fu Yeh, Prakash Murugesan, Peyman Heidari, Yue Liu, Kavya Srinet, Babak Damavandi and Anuj Kumar	1314
<i>SLM as Guardian: Pioneering AI Safety with Small Language Model</i> Ohjoon Kwon, Donghyeon Jeon, Nayoung Choi, Gyu-Hwung Cho, Hwiyeol Jo, Changbong Kim, Hyunwoo Lee, Inho Kang, Sun Kim and Taiwoo Park	1333
<i>Hyper-QKSG: Framework for Automating Query Generation and Knowledge-Snippet Extraction from Tables and Lists</i> Dooyoung Kim, Yoonjin Jang, Dongwook Shin, Chanhoon Park and Youngjoong Ko	1351
<i>Patentformer: A Novel Method to Automate the Generation of Patent Applications</i> Juanyan Wang, Sai Krishna Reddy Mudhiganti and Manali Sharma	1361

<i>MARCO: Multi-Agent Real-time Chat Orchestration</i>	
Anubhav Shrimal, Stanley Kanagaraj, Kriti Biswas, Swarnalatha Raghuraman, Anish Nediyan- chath, Yi Zhang and Promod Yenigalla	1381
<i>mGTE: Generalized Long-Context Text Representation and Reranking Models for Multilingual Text Retrieval</i>	
Xin Zhang, Yanzhao Zhang, Dingkun Long, Wen Xie, Ziqi Dai, Jialong Tang, Huan Lin, Baosong Yang, Pengjun Xie, Fei Huang, Meishan Zhang, Wenjie Li and Min Zhang	1393
<i>ItiNera: Integrating Spatial Optimization with Large Language Models for Open-domain Urban Itine- rary Planning</i>	
Yihong Tang, Zhaokai Wang, Ao Qu, Yihao Yan, Zhaofeng Wu, Dingyi Zhuang, Jushi Kai, Ke- bing Hou, Xiaotong Guo, Jinhua Zhao, Zhan Zhao and Wei Ma	1413
<i>RESTful-Llama: Connecting User Queries to RESTful APIs</i>	
Han Xu, Ruining Zhao, Jindong Wang and Haipeng Chen	1433
<i>A Cost-Efficient Modular Sieve for Extracting Product Information from Company Websites</i>	
Anna Hättö, Dragan Milchevski, Kersten Döring, Marko Putnikovic, Mohsen Mesgar, Filip No- vović, Maximilian Braun, Karina Leoni Borimann and Igor Stranjanac	1444
<i>CharacterGLM: Customizing Social Characters with Large Language Models</i>	
Jinfeng Zhou, Zhuang Chen, Dazhen Wan, Bosi Wen, Yi Song, Jifan Yu, Yongkang Huang, Pei Ke, Guanqun Bi, Libiao Peng, JiaMing Yang, Xiyao Xiao, Sahand Sabour, Xiaohan Zhang, Wenjing Hou, Yijia Zhang, Yuxiao Dong, Hongning Wang, Jie Tang and Minlie Huang	1457
<i>RAC: Retrieval-augmented Conversation Dataset for Open-domain Question Answering in Conversa- tional Settings</i>	
Bonggeun Choi, JeongJae Park, Yoonsung Kim, Jaehyun Park and Youngjoong Ko	1477
<i>Improving Retrieval in Sponsored Search by Leveraging Query Context Signals</i>	
Akash Kumar Mohankumar, Gururaj K, Gagan Madan and Amit Singh	1489
<i>FuxiTranyu: A Multilingual Large Language Model Trained with Balanced Data</i>	
Haoran Sun, Renren Jin, Shaoyang Xu, Leiyu Pan, Supryadi , Menglong Cui, Jiangcun Du, Yikun Lei, Lei Yang, Ling Shi, Juesi Xiao, Shaolin Zhu and Deyi Xiong	1499
<i>QUIS: Question-guided Insights Generation for Automated Exploratory Data Analysis</i>	
Abhijit Manatkar, Ashlesha Akella, Parthivi Gupta and Krishnasuri Narayanam	1523
<i>PEARL: Preference Extraction with Exemplar Augmentation and Retrieval with LLM Agents</i>	
Vijit Malik, Akshay Jagatap, Vinayak S Puranik and Anirban Majumder	1536
<i>RAG-HAT: A Hallucination-Aware Tuning Pipeline for LLM in Retrieval-Augmented Generation</i>	
Juntong Song, Xingguang Wang, Juno Zhu, Yuanhao Wu, Xuxin Cheng, Randy Zhong and Cheng Niu	1548
<i>Intent Detection in the Age of LLMs</i>	
Gaurav Arora, Shreya Jain and Srujana Merugu	1559
<i>Aegis: An Advanced LLM-Based Multi-Agent for Intelligent Functional Safety Engineering</i>	
Lu Shi, Bin Qi, Jiarui Luo, Yang Zhang, Zhanzhao Liang, Zhaowei Gao, Wenke Deng and Lin Sun	1571
<i>Efficient Answer Retrieval System (EARS): Combining Local DB Search and Web Search for Generative QA</i>	
Nikita Krayko, Ivan Sidorov, Fedor Laputin, Daria Galimzianova and Vasily Kononov	1584

<i>GraphQL Query Generation: A Large Training and Benchmarking Dataset</i>	
Manish Kesarwani, Sambit Ghosh, Nitin Gupta, Shramona Chakraborty, Renuka Sindhgatta, Sameep Mehta, Carlos Eberhardt and Dan Debrunner	1595
<i>Mixture of Diverse Size Experts</i>	
Manxi Sun, Wei Liu, Jian Luan, Pengzhi Gao and Bin Wang	1608
<i>Course-Correction: Safety Alignment Using Synthetic Preferences</i>	
Rongwu Xu, Yishuo Cai, Zhenhong Zhou, Renjie Gu, Haiqin Weng, Liu Yan, Tianwei Zhang, Wei Xu and Han Qiu	1622
<i>GOVERN: Gradient Orientation Vote Ensemble for Multi-Teacher Reinforced Distillation</i>	
Wenjie Zhou, Zhenxin Ding, Xiaodong Zhang, Haibo Shi, Junfeng Wang and Dawei Yin...	1650
<i>PRISM: A New Lens for Improved Color Understanding</i>	
Arjun Reddy Akula, Garima Pruthi, Inderjit S Dhillon, Pradyumna Narayana, Sugato Basu and Varun Jampani	1659