

NAACL 2025

**Annual Conference of the Nations of the Americas Chapter of
the Association for Computational Linguistics**

**Proceedings of the Conference
Industry Track**

April 30, 2025

©2025 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-194-0

Table of Contents

<i>Understanding LLM Development Through Longitudinal Study: Insights from the Open Ko-LLM Leaderboard</i>	
Chanjun Park and Hyeonwoo Kim	1
<i>RTSM: Knowledge Distillation with Diverse Signals for Efficient Real-Time Semantic Matching in E-Commerce</i>	
Sanjay Agrawal and Vivek Sembium	9
<i>WorkTeam: Constructing Workflows from Natural Language with Multi-Agents</i>	
Hanchao Liu, Rongjun Li, Weimin Xiong, Ziyu Zhou and Wei Peng	20
<i>How LLMs React to Industrial Spatio-Temporal Data? Assessing Hallucination with a Novel Traffic Incident Benchmark Dataset</i>	
Qiang Li, Mingkun Tan, Xun Zhao, Dan Zhang, Daoan Zhang, Shengzhao Lei, Anderson S. Chu, Lujun Li and Porawit Kamnoedboon	36
<i>Text2Sql: Pure Fine-Tuning and Pure Knowledge Distillation</i>	
Gao yu Zhu, Wei Shao, Xichou Zhu, Lei Yu, Jiafeng Guo and Xueqi Cheng	54
<i>MoEMoE: Question Guided Dense and Scalable Sparse Mixture-of-Expert for Multi-source Multi-modal Answering</i>	
Vinay Kumar Verma, Shreyas Sunil Kulkarni, Happy Mittal and Deepak Gupta	62
<i>Finding-Centric Structuring of Japanese Radiology Reports and Analysis of Performance Gaps for Multiple Facilities</i>	
Yuki Tagawa, Yohei Momoki, Norihisa Nakano, Ryota Ozaki, Motoki Taniguchi, Masatoshi Hori and Noriyuki Tomiyama	70
<i>Learning LLM Preference over Intra-Dialogue Pairs: A Framework for Utterance-level Understandings</i>	
Xuanqing Liu, Luyang Kong, Wei Niu, Afshin Khashei, Belinda Zeng, Steve Johnson, Jon Jay, Davor Golac and Matt Pope	86
<i>Enhancing Function-Calling Capabilities in LLMs: Strategies for Prompt Formats, Data Integration, and Multilingual Translation</i>	
Yi-Chang Chen, Po-Chun Hsu, Chan-Jan Hsu and Da-shan Shiu	99
<i>Exploring Straightforward Methods for Automatic Conversational Red-Teaming</i>	
George Kour, Naama Zwerdling, Marcel Zalmanovici, Ateret Anaby Tavor, Ora Nova Fandina and Eitan Farchi	112
<i>A Diverse and Effective Retrieval-Based Debt Collection System with Expert Knowledge</i>	
Jiaming Luo, Weiyi Luo, Guoqing Sun, Mengchen Zhu, Haifeng Tang, Kenny Q. Zhu and Mengyue Wu	129
<i>Search Query Embeddings via User-behavior-driven Contrastive Learning</i>	
Sosuke Nishikawa, Jun Hirako, Nobuhiro Kaji, Koki Watanabe, Hiroki Asano, Souta Yamashiro and Shumpei Sano	138
<i>QSpell 250K: A Large-Scale, Practical Dataset for Chinese Search Query Spell Correction</i>	
Dezhi Ye, Haomei Jia, Junwei Hu, Tian Bowen, Jie Liu, Haijin Liang, Jin Ma and Wenmin Wang	
148	

<i>CONSTRUCTA: Automating Commercial Construction Schedules in Fabrication Facilities with Large Language Models</i>	
Yifan Zhang and Xue Yang	156
<i>Challenges and Remedies of Domain-Specific Classifiers as LLM Guardrails: Self-Harm as a Case Study</i>	
Bing Zhang and Guang-Jie Ren	173
<i>Mitigating Bias in Item Retrieval for Enhancing Exam Assembly in Vocational Education Services</i>	
Alonso Palomino, Andreas Fischer, David Buschhüter, Roland Roller, Niels Pinkwart and Benjamin Paassen	183
<i>Breaking Boundaries: Investigating the Effects of Model Editing on Cross-linguistic Performance</i>	
Somnath Banerjee, Avik Halder, Rajarshi Mandal, Sayan Layek, Ian Soboroff, Rima Hazra and Animesh Mukherjee	194
<i>Towards Reliable and Practical Phishing Detection</i>	
Hyowon Cho and Minjoon Seo	210
<i>Zero-Shot ATC Coding with Large Language Models for Clinical Assessments</i>	
Zijian Chen, John-Michael Gamble and Jimmy Lin	226
<i>Navigating the Path of Writing: Outline-guided Text Generation with Large Language Models</i>	
Yukyung Lee, Soonwon Ka, Bokyung Son, Pilsung Kang and Jaewook Kang	233
<i>TaeBench: Improving Quality of Toxic Adversarial Examples</i>	
Jennifer Zhu, Dmitriy Bespalov, Liwen You, Ninad Kulkarni and Yanjun Qi	251
<i>Open Ko-LLM Leaderboard2: Bridging Foundational and Practical Evaluation for Korean LLMs</i>	
Hyeonwoo Kim, Dahyun Kim, Jihoo Kim, Sukyung Lee, Yungi Kim and Chanjun Park	266
<i>CuriousLLM: Elevating Multi-Document Question Answering with LLM-Enhanced Knowledge Graph Reasoning</i>	
Zukang Yang, Zixuan Zhu and Jennifer Zhu	274
<i>CharacterGPT: A Persona Reconstruction Framework for Role-Playing Agents</i>	
Jeiyoon Park, Chanjun Park and Heuiseok Lim	287
<i>Efficient Continual Pre-training of LLMs for Low-resource Languages</i>	
Arijit Nag, Soumen Chakrabarti, Animesh Mukherjee and Niloy Ganguly	304
<i>DSRAG: A Double-Stream Retrieval-Augmented Generation Framework for Countless Intent Detection</i>	
Pei Guo, Enjie Liu, Ruichao Zhong, Mochi Gao, Yunzhi Tan, Bo Hu and Zang Li	318
<i>Octopus: On-device language model for function calling of software APIs</i>	
Wei Chen, Zhiyuan Li and Mingyuan MA	329
<i>MoFE: Mixture of Frozen Experts Architecture</i>	
Jean Seo, Jaeyoon Kim and Hyopil Shin	340
<i>FinLLM-B: When Large Language Models Meet Financial Breakout Trading</i>	
Kang Zhang, Osamu Yoshie, Lichao Sun and Weiran Huang	349
<i>QueryShield: A Platform to Mitigate Enterprise Data Leakage in Queries to External LLMs</i>	
Nitin Ramrakhiani, Delton Myalil, Sachin Pawar, Manoj Apte, Rajan M A, Divyesh Saglani and Imtiyazuddin Shaik	358

<i>SwissADT: An Audio Description Translation System for Swiss Languages</i>	
Lukas Fischer, Yingqiang Gao, Alexa Lintner, Annette Rios and Sarah Ebling	370
<i>Chinese Morph Resolution in E-commerce Live Streaming Scenarios</i>	
Jiahao Zhu, Jipeng Qiang, Ran Bai, Chenyu Liu and Xiaoye Ouyang	380
<i>MonoTODia: Translating Monologue Requests to Task-Oriented Dialogues</i>	
Sebastian Steindl, Ulrich Schäfer and Bernd Ludwig	390
<i>MedEthicEval: Evaluating Large Language Models Based on Chinese Medical Ethics</i>	
Haoan Jin, Jiacheng Shi, Hanhui Xu, Kenny Q. Zhu and Mengyue Wu	404
<i>Predicting ICU Length of Stay for Patients using Latent Categorization of Health Conditions</i>	
Tirthankar Dasgupta, Manjira Sinha and Sudeshna Jana	422
<i>RevieWeaver: Weaving Together Review Insights by Leveraging LLMs and Semantic Similarity</i>	
Jiban Adhikary, Mohammad Alqudah and Arun Palghat Udayashankar	431
<i>MedCodER: A Generative AI Assistant for Medical Coding</i>	
Krishanu Das Baksi, Elijah Soba, John J Higgins, Ravi Saini, Jaden Wood, Jane Cook, Jack I Scott, Nirmala Pudota, Tim Weninger, Edward Bowen and Sanmitra Bhattacharya	449
<i>Visual Zero-Shot E-Commerce Product Attribute Value Extraction</i>	
Jiaying Gong, Ming Cheng, Hongda Shen, Pierre-Yves Vandenbussche, Janet Jenq and Hoda Eldardiry	460
<i>SCORE: Systematic Consistency and Robustness Evaluation for Large Language Models</i>	
Grigor Nalbandyan, Rima Shahbazyan and Evelina Bakhturina	470
<i>Evaluating Large Language Models with Enterprise Benchmarks</i>	
Bing Zhang, Mikio Takeuchi, Ryo Kawahara, Shubhi Asthana, Maruf Hossain, Guang-Jie Ren, Kate Soule, Yifan Mai and Yada Zhu	485
<i>Can Post-Training Quantization Benefit from an Additional QLoRA Integration?</i>	
Xiliang Zhu, Elena Khasanova and Cheng Chen	506
<i>From Generating Answers to Building Explanations: Integrating Multi-Round RAG and Causal Modeling for Scientific QA</i>	
Victor Barres, Clifton James McFate, Aditya Kalyanpur, Kailash Karthik Saravanakumar, Lori Moon, Natnael Seifu and Abraham Bautista-Castillo	515
<i>TurboFuzzLLM: Turbocharging Mutation-based Fuzzing for Effectively Jailbreaking Large Language Models in Practice</i>	
Aman Goel, Xian Wu, Zhe Wang, Dmitriy Besspalov and Yanjun Qi	523
<i>Does Self-Attention Need Separate Weights in Transformers?</i>	
Md Kowsher, Nusrat Jahan Prottasha, Chun-Nam Yu, Ozlem Garibay and Niloofar Yousefi	535
<i>SuperRAG: Beyond RAG with Layout-Aware Graph Modeling</i>	
Chening Yang, Duy-Khanh Vu, Minh-Tien Nguyen, Xuan-Quang Nguyen, Linh Nguyen and Hung Le	544
<i>SweEval: Do LLMs Really Swear? A Safety Benchmark for Testing Limits for Enterprise Use</i>	
Hitesh Laxmichand Patel, Amit Agarwal, Arion Das, Bhargava Kumar, Srikant Panda, Priyaranjan Pattanayak, Taki Hasan Rafi, Tejaswini Kumar and Dong-Kyu Chae	558
<i>Natural Language Processing for Human Resources: A Survey</i>	
Naoki Otani, Nikita Bhutani and Estevam Hruschka	583

<i>Implementing Retrieval Augmented Generation Technique on Unstructured and Structured Data Sources in a Call Center of a Large Financial Institution</i>	
Syed Shariyar Murtaza, Yifan Nie, Elias Avan, Utkarsh Soni, Wanyu Liao, Adam Carnegie, Cyril John Mathias, Junlin Jiang and Eugene Wen	598
<i>Granite Guardian: Comprehensive LLM Safeguarding</i>	
Inkit Padhi, Manish Nagireddy, Giandomenico Cornacchia, Subhajit Chaudhury, Tejaswini Pedapati, Pierre Dognin, Keerthiram Murugesan, Erik Miebling, Martín Santillán Cooper, Kieran Fraser, Giulio Zizzo, Muhammad Zaid Hameed, Mark Purcell, Michael Desmond, Qian Pan, Inge Vejsbjerg, Elizabeth M. Daly, Michael Hind, Werner Geyer, Ambrish Rawat, Kush R. Varshney and Prasanna Sattigeri	607
<i>Breaking Down Power Barriers in On-Device Streaming ASR: Insights and Solutions</i>	
Yang Li, Yuan Shanguan, Yuhao Wang, Liangzhen Lai, Ernie Chang, Changsheng Zhao, Yangyang Shi and Vikas Chandra	616
<i>Break-Ideate-Generate (BRIDGE): Moving beyond Translations for Localization using LLMs</i>	
Swapnil Gupta, Lucas Pereira Carlini, Prateek Sircar and Deepak Gupta	627
<i>Concept Distillation from Strong to Weak Models via Hypotheses-to-Theories Prompting</i>	
Emmanuel Aboah Boateng, Cassiano O Becker, Nabiha Asghar, Kabir Walia, Ashwin Srinivasan, Ehi Nosakhare, Soundararajan Srinivasan and Victor Dibia	638
<i>Towards Reliable Agents: Benchmarking Customized LLM-Based Retrieval-Augmented Generation Frameworks with Deployment Validation</i>	
Kevin Shukang Wang, Karel Joshua Harjono and Ramon Lawrence	655
<i>Query Variant Detection Using Retriever as Environment</i>	
Minji Seo, Youngwon Lee, Seung-won Hwang, Seoho Song, Hee-Cheol Seo and Young-In Song	662
<i>Evaluating Bias in LLMs for Job-Resume Matching: Gender, Race, and Education</i>	
Hayate Iso, Pouya Pezeshkpour, Nikita Bhutani and Estevam Hruschka	672
<i>Goal-Driven Data Story, Narrations and Explanations</i>	
Aniya Aggarwal, Ankush Gupta, Shivangi Bithel and Arvind Agarwal	684
<i>VIT-Pro: Visual Instruction Tuning for Product Images</i>	
Vishnu Prabhakaran, Purav Aggarwal, Vishruit Kulshreshtha, Arunita Das, Sahini Venkata Sitaran Sruti and Anoop Saladi	695
<i>AutoKB: Automated Creation of Structured Knowledge Bases for Domain-Specific Support</i>	
Rishav Sahay, Arihant Jain, Purav Aggarwal and Anoop Saladi	708
<i>Medical Spoken Named Entity Recognition</i>	
Khai Le-Duc, David Thulke, Hung-Phong Tran, Long Vo-Dang, Khai-Nguyen Nguyen, Truong-Son Hy and Ralf Schlüter	724
<i>PLEX: Adaptive Parameter-Efficient Fine-Tuning for Code LLMs using Lottery-Tickets</i>	
Jaeseong Lee, Hojae Han, Jongyoon Kim, Seung-won Hwang, Naun Kang, KyungJun An and Sungho Jang	784
<i>Evaluating the Performance of RAG Methods for Conversational AI in the Airport Domain</i>	
Yuyang Li, Pjm Kerbusch, Rhr Pruum and Tobias Käfer	794
<i>LLM Safety for Children</i>	
Prasanjit Rath, Hari Shrawgi, Parag Agrawal and Sandipan Dandapat	809

<i>RxLens: Multi-Agent LLM-powered Scan and Order for Pharmacy</i> Akshay Jagatap, Srujana Merugu and Prakash Mandayam Comar	822
<i>Distill-C: Enhanced NL2SQL via Distilled Customization with LLMs</i> Cong Duy Vu Hoang, Gioacchino Tangari, Clemence Lanfranchi, Dalu Guo, Paul Cayet, Steve Siu, Don Dharmasiri, Yuan-Fang Li, Long Duong, Damien Hilloulin, Rhicheck Patra, Sungpack Hong and Hassan Chafi	833
<i>eC-Tab2Text: Aspect-Based Text Generation from e-Commerce Product Tables</i> Luis Antonio Gutierrez Guanilo, Mir Tafseer Nayeem, Cristian Jose Lopez Del Alamo and Davood Rafiei	849
<i>RAD-Bench: Evaluating Large Language Models' Capabilities in Retrieval Augmented Dialogues</i> Tzu-Lin Kuo, FengTing Liao, Mu-Wei Hsieh, Fu-Chieh Chang, Po-Chun Hsu and Da-shan Shiu	868
<i>Conflict and Overlap Classification in Construction Standards Using a Large Language Model</i> Seong-Jin Park, Youn-Gyu Jin, Hyun-Young Moon, Choi Bong-Hyuck, Lee Seung Hwan, Ohjoon Kwon and Kang-Min Kim	903
<i>Protein2Text: Resampling Mechanism to Translate Protein Sequences into Human-Interpretable Text</i> Ala Jararweh, Oladimeji Macaulay, David Arredondo, Yue Hu, Luis E Tafoya, Kushal Virupakshappa and Avinash Sahu	918
<i>Cracking the Code: Multi-domain LLM Evaluation on Real-World Professional Exams in Indonesia</i> Fajri Koto	938
<i>CodeGenWrangler: Data Wrangling task automation using Code-Generating Models</i> Ashlesha Akella, Abhijit Manatkar, Krishnasuri Narayanam and Sameep Mehta	949
<i>Dialogue Language Model with Large-Scale Persona Data Engineering</i> Mengze Hong, Chen Jason Zhang, Chaotao Chen, Rongzhong Lian and Di Jiang	961
<i>Developing a Reliable, Fast, General-Purpose Hallucination Detection and Mitigation Service</i> Song Wang, Xun Wang, Jie Mei, Yujia Xie, Si-Qing Chen and Wayne Xiong	971
<i>Improved Near-Duplicate Detection for Aggregated and Paywalled News-Feeds</i> Siddharth Tumre, Sangameshwar Patil and Alok Kumar	979
<i>Pisets: A Robust Speech Recognition System for Lectures and Interviews</i> Ivan Bondarenko, Daniil Grebenkin, Oleg Sedukhin, Mikhail Klementev, Derunets Roman and Lyudmila Budneva	988
<i>CPRM: A LLM-based Continual Pre-training Framework for Relevance Modeling in Commercial Search</i> Kaixin Wu, Yixin Ji, Zeyuan Chen, Qiang Wang, Cunxiang Wang, Hong Liu, Baijun Ji, Xu Jia, Zhongyi Liu, Jinjie GU, Yuan Zhou and Linjian Mo	998
<i>Schema and Natural Language Aware In-Context Learning for Improved GraphQL Query Generation</i> Nitin Gupta, Manish Kesarwani, Sambit Ghosh, Sameep Mehta, Carlos Eberhardt and Dan Debrunner	1009
<i>Chatbot Arena Estimate: towards a generalized performance benchmark for LLM capabilities</i> Lucas Spangher, Tianle Li, William F. Arnold, Nick Masiewicki, Xerxes Dotiwalla, Rama Kumar Pasumarthi, Peter Grabowski, Eugene Ie and Daniel Gruhl	1016

<i>Enhancing Temporal Understanding in Audio Question Answering for Large Audio Language Models</i> Arvind Krishna Sridhar, Yinyi Guo and Erik Visser	1026
<i>HyPA-RAG: A Hybrid Parameter Adaptive Retrieval-Augmented Generation System for AI Legal and Policy Applications</i> Rishi Kalra, Zekun Wu, Ayesha Gulley, Airlie Hilliard, Xin Guan, Adriano Koshiyama and Philip Colin Treleaven	1036
<i>An Efficient Context-Dependent Memory Framework for LLM-Centric Agents</i> Pengyu Gao, Jinming Zhao, Xinyue Chen and Long Yilin	1055