

EMNLP 2026

**The 2026 Conference on Empirical Methods in Natural
Language Processing**

Industry Track

October 24-29, 2026

The EMNLP organizers gratefully acknowledge the support from the following sponsors.

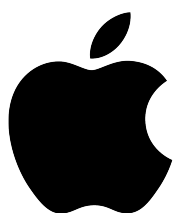
Diamond

Bloomberg

Engineering

Google

Platinum



Baidu 百度

Capital One

cohere

GPTZero



Megagon Labs

ORACLE

rednote

airbnb

Gold

蚂蚁集团
ANT GROUP

JPMorganChase

美团 Meituan

SAP

Tencent

tether.

Silver

Adobe

Bronze

DATAmundi.ai

DataoceanAI

t translated.

Booking.com

ALICE

Keenable


**Thomson
Reuters™**

 **oruk**

Academic Ally



©2026 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 979-8-89176-460-6

Industry Track Preface

Welcome to the EMNLP Industry Track 2026!

The EMNLP 2026 Industry Track highlights insights, research challenges, and practical lessons arising from the design, development, deployment, and evaluation of real-world language and speech technology applications. The track welcomes work on system design, efficiency, maintainability and scalability, novel applications and use cases, and methods for deployed systems, with the broader goal of strengthening the exchange of ideas and experience between academia and industry.

The Industry Track was first introduced at EMNLP in 2022, and EMNLP 2026 marks its fifth edition at the conference. This year, the track received 931 submissions. Submissions were reviewed in a double-blind process and assessed based on novelty, technical quality, potential impact, and clarity, with particular emphasis on real-world implementations, real-world data, and appropriate evaluation methodology. Authors were given an opportunity to respond to reviews before final decisions were made.

Following the review process, 265 papers were selected for presentation at EMNLP 2026. Of these, 23 were selected for oral presentation and 242 for poster presentation. The Industry Track programme will take place in Budapest, Hungary, on October 25–27, 2026, as part of EMNLP 2026.

We would like to thank all authors who submitted their work to the Industry Track, as well as the reviewers and area chairs who contributed their time and expertise to the review process and helped shape the final program. We are also grateful to the EMNLP 2026 organizing committee and to the chairs of previous Industry Tracks for their guidance and support.

We hope you find the proceedings useful and enjoyable.

Enjoy the conference!

The EMNLP 2026 Industry Track Programme Co-Chairs

Organizing Committee

Industry Track Chairs

Pepa Atanasova, University of Copenhagen

Gholamreza Haffari, Monash University

Minghan Wang, Monash University

Program Committee

Program Chairs

Pepa Atanasova, University of Copenhagen
Gholamreza Haffari, Monash University
Minghan Wang, Monash University

Area Chairs

Zubair Afzal, Elsevier
Amit Agarwal, Oracle
Firoj Alam, Qatar Computing Research Institute
Aijun An, York University
Pepa Atanasova, University of Copenhagen
Lei Bai, Shanghai AI Laboratory
Mariam Barry
Farah Benamara, Institut de recherche en informatique de toulouse
Abdallah Benzine
Johannes Bjerva, Aalborg University
Nicolas Bougie
Quentin Brabant, Orange-labs
Evgeny Burnaev, Applied AI Institute
Liangliang Cao, Google
Vittorio Castelli, Oracle
Chung-Chi Chen, NII, Tokyo Institute of Technology
Pin-Yu Chen, International Business Machines
Yimin Chen, University of Massachusetts at Lowell
Zulong Chen
Chanyeol Choi
Jaegul Choo, Korea Advanced Institute of Science and Technology
Prafulla Kumar Choubey, Salesforce.com
Sagnik Ray Choudhury, University of North Texas
Daniel Dahlmeier, Heidelberg University, Ruprecht-Karls-Universität Heidelberg, Singapore University of Technology and Design and SAP SE
Yang Deng, Singapore Management University
Dimitrios Dimitriadis, Amazon
Xiaoyi Dong, Microsoft
Pablo Duboue, Internet Archive and Textualization Software Ltd.
Ondrej Dusek, Charles University, Prague
Christos Emmanouilidis, University of Groningen and Athena Research and Innovation Centre
Fuli Feng, University of Science and Technology of China
Jie Feng, Zhongguancun Academy
Xingcheng Fu
Xue-Yong Fu
Chen Gao, Tsinghua University
Lingyu Gao, Duolingo and Circle Cat Inc
Kallirroi Georgila, University of Southern California and USC/ISI
Chenlei Guo, Amazon
Gholamreza Haffari, Monash University

Fengxiang He, University of Edinburgh
 Wei Hu, Nanjing University
 Xuming Hu, The Hong Kong University of Science and Technology (Guangzhou) and Hong Kong
 University of Science and Technology
 Md Mofijul Islam, Amazon
 Beakcheol Jang, Yonsei University
 Luo Ji, Beijing Foreign Studies University
 Wenhao Jiang, Guangming Laboratory
 Baoyu Jing, Meta
 Mohammad Kachuee, Oracle
 Evgeny Kharlamov, University of Oslo and Robert Bosch GmbH, Bosch
 Takeshi Kojima, The University of Tokyo
 Adit Krishnan, Amazon
 Lun-Wei Ku, Academia Sinica
 Bhargava Kumar, TD Securities
 Branislav Kveton, Adobe Research
 Kwei-Herng Lai, Microsoft
 Jean-Charles Lamirel, University of Pantheon Sorbonne, Dalian University of Technology, Univer-
 sité de Strasbourg, University of Lorraine (LORIA) and University of SStrasbourg
 Yunshi Lan, East China Normal University
 Md Tahmid Rahman Laskar, York University and Dialpad Inc.
 Jongwuk Lee, Sungkyunkwan University
 Yongjae Lee, LinqAlpha and Ulsan National Institute of Science and Technology
 Pengyuan Li, IBM Research
 Jianzhe Lin
 Weizhe Lin, Huawei Technologies Ltd.
 Yue Liu, National University of Singapore
 Zechun Liu, Meta Inc.
 Yunfei Long, Queen Mary, University of London
 Ziyang Luo, Amazon
 Qiyao Ma, University of California, Davis
 Marco Maggini
 Shervin Malmasi, Amazon
 Sarah Masud
 Yashar Mehdad, Airbnb
 Bonan Min, Airbnb
 Bamdev Mishra, Microsoft
 Nabeel Mohammed, North South University
 Will Monroe, Duolingo
 Sebastien Montella, Huawei Technologies R&D UK Ltd
 Alessandro Moschitti, Amazon
 Pradyumna Narayana, Google
 Usman Naseem, Macquarie University
 Patrick Ng, Cornell University and Oracle
 Huy V. Nguyen, Oracle
 Sarana Nutanong
 Slim Ouni, Université de Lorraine
 Inkit Padhi
 Chanjun Park, Soongsil University
 Cheonbok Park, Korea Advanced Institute of Science & Technology and NAVER
 Yannick Parmentier, Université de Lorraine

Dhaval C Patel, International Business Machines
 Hitesh Laxmichand Patel, Oracle
 Yifan Peng, Weill Cornell Medicine, Cornell University
 Tatiana Petrova, University of Luxemburg
 Francesco Pierri, Politecnico di Milano
 Aske Plaat, Leiden University
 Arantza del Pozo, Fundación Vicomtech
 Fanchao Qi
 Chen Qian, Shanghai Jiaotong University
 Xiaoting Qin, Microsoft
 Bhaktipriya Radharapu, Facebook
 Ambrish Rawat, International Business Machines
 Shaina Raza, Vector institute
 Tong Ruan
 Anit Kumar Sahu, Oracle
 Matthew Shardlow, The Manchester Metropolitan University
 Artem Shelmanov, Mohamed bin Zayed University of Artificial Intelligence
 Xiaoyu Shen, Eastern Institute of Technology, Ningbo
 Ashish Shenoy, Meta
 Manish Shrivastava, International Institute of Information Technology Hyderabad
 Kevin Small, Microsoft
 Charese Smiley, J.P. Morgan Chase
 Andy Song, RMIT University
 Yang Song, Facebook
 Radu State, University of Luxemburg
 Kai Sun, Meta
 Zequn Sun, Nanjing University
 Zhoujian Sun
 Martin Takáč, Mohamed bin Zayed University of Artificial Intelligence
 Zhi Rui Tam, National Taiwan University and Appier
 Changxin Tian, Ant Group
 Zhaopeng Tu, Tencent
 Chengyu Wang, Independent Researcher and Alibaba Group
 Haixun Wang, AI
 Hongwei Wang, Meta
 Junjie Wang, Tsinghua University
 Shuaiqiang Wang, Baidu Inc.
 Wenjie Wang, University of Science and Technology of China
 Taro Watanabe, Nara Institute of Science and Technology, Japan
 Zeyi Wen, Hong Kong University of Science and Technology (Guangzhou)
 Genta Indra Winata, Capital One
 Fanyou Wu, Amazon
 Lijun Wu, Shanghai Artificial Intelligence Laboratory
 Liwei WU, LinkedIn and LinkedIn
 Dongbo Xi
 Zhen Xiang, University of Georgia
 Tong Xiao, Northeastern University
 Xike Xie
 Deyi Xiong, Tianjin University
 Lingyong Yan, Baidu Inc.
 Chao-Han Huck Yang, Apple

Fan Yang, Wake Forest University
 Tong Yang, Peking University
 Yujiu Yang, Tsinghua University
 Liang Yao, SUN YAT-SEN UNIVERSITY
 SM Yiu, University of Hong Kong
 Kang Min Yoo, Amazon
 Seunghyun Yoon, Adobe Research
 Yuhang Zang, Shanghai Artificial Intelligence Laboratory
 Bo Zhang, Agent Center, Shanghai AI Laboratory
 Hainan Zhang
 Jiaxin Zhang, Salesforce AI Research
 Kaiyan Zhang, Frontis.AI
 Linfeng Zhang, Shanghai Jiaotong University
 Wei Zhang, Eastern Institute of Technology, Ningbo, China.
 Wentao Zhang, Peking University
 Yongqi Zhang, The Hong Kong University of Science and Technology (Guangzhou)
 Yilun Zhao, Yale University
 Xiatian Zhu, University of Surrey
 Fuzhen Zhuang, School of Artificial Intelligence, Beihang University and Institute of Computing Technology, Chinese Academy of Sciences

Reviewers

Ahmed Abdelali, Hanna Abi Akl, Mohammad Yeghaneh Abkenar, Satyen Abrol, Rupam Acharyya, Sharad Agarwal, Ayushi Aggarwal, Aastha Agrawal, Aditya Agrawal, Samarth Agrawal, Sanjay Agrawal, Mazen Mohmed Nasr Mohamed Ahmed, Zeeshan Ahmed, Sarthak Ahuja, Shayan Ali Akbar, Arda Akdemir, Ömer Faruk Akgül, Shahnawaz Alam, Georgios Alexandridis, Rafiq Ali, Héctor Allende-Cid, Omar Alonso, Laura Alonso Alemany, Duygu Altinok, Mayank Anand, Raviteja Anantha, Rafael Anchiêta, Enrique Henestroza Anguiano, Ameya Anjarlekar, Negar Arabzadeh, Yokila Arora, Karthika Arumugam, Abhinav Arun, Ankit Arun, Shubhi Asthana, Àlex R. Atrio, Joseph Axisa, Mohammad Anas Azeez

Sanket Badhe, Ahsaas Bajaj, Mithun Balakrishna, Kartik Balasubramaniam, Ankita Banerjee, Yuwei Bao, Roberto Barra-Chicote, Leslie Barrett, Lee Becker, Radhika Bhargava, Hanoz Bhatena, Gagan Bhatia, Nidhi Bhatia, Sudhanshu Bhoi, Nikita Bhutani, Zhen Bi, Debmalaya Biswas, Umesh Bodhwani, Bogdan Bogachov, Mihaela Bornea, Nadjat Bouayad-Agha, Nicolas Bougie, Antoine Boutet, Leonid Boytsov, Thomas Brovelli, Marco Andrea Buchmann, Sebastiaan Buiting, Paweł Bujnowski

Yu-Tong Cao, Fabio Casati, Chompakorn Chaksangchaichot, Nischal Reddy Chandra, Kaiyan Chang, Ankani Chatteraj, Baiyu Chen, Chien-Hung Chen, Fuxiang Chen, Guanhua Chen, Jiangning Chen, Jingquan Chen, Jinwen Chen, John Chen, Junwen Chen, Mengxiang Chen, Qinglin Chen, You-Lin Chen, Yuan Chen, Yubo Chen, Yuxu Chen, Dawei Cheng, Song Cheng, Yexiang Cheng, Maitreya Prafulla Chitale, Eunah Cho, Won Ik Cho, Youngwon Choi, Shamil Chollampatt, Pei Chu, Yun-Shiuan Chuang, Tarang Chugh, Sejin Chun, Pedro Cisneros-Velarde, Bruno Gomes Coelho, Giandomenico Cornacchia, Adrian Cosma, Wendi Cui, Osman Alperen Çinar-Koraş

Deborah A. Dahl, Daniel Dakota, Arup Kumar Das, Devleena Das, Sambit Das, Souvik Das, Steve DeNeefe, Daryna Dementieva, Cheng Deng, Wenlong Deng, Harsh Rao Dhanyamraju, Prajit Dhar, Peng Di, Maryam Dialameh, Rogelio Diaz-Mendez, Dimitar Iliyanov Dimitrov, Hailun Ding, Weihang Ding, Rahul Divekar, Mohit Diwan, Bin Dong, Anuj Doshi, P. Alex Dow, Pan

Du, Karan Dua, Juanyong Duan, Matthew T. Dunn, Srimonti Dutta, Guy Stéphane Waffo Dzuyo, Altigran da Silva, Edleno Silva de Moura

Keelan Evanini

Run-Ze Fan, Tao Fan, Maryam Fatima, Huijie Feng, Michael Flor, Nicolas Flores-Herr, Simona Frenda, Lisheng Fu

Charan S G, Madhava Gaikwad, Ankur Gandhe, Chongyang Gao, Jiechao Gao, Jiechao Gao, Radhika Gaonkar, Aashna Garg, Yirou Ge, Advitya Gemawat, Hejia Geng, Rajat Ghosh, Sucheta Ghosh, Kellen Gillespie, Voula Giouli, Bogdan Gliwa, Anmol Goel, Vinod Goje, Noon Pokarat-siri Goldstein, Andrey Goncharov, Jiaying Gong, Madhu Gopinathan, Janaki Ram Goteti, Konstantin Grotov, Mao Guan, Sheng Guan, Gaël Guibon, Arush Gumber, Dongxin Guo, Xiaoyu Guo, Zhongbin Guo, Ashray Gupta, Deepak Gupta, Manish Gupta, Sachin Gupta, Sameer Gupta, Vipul Gupta, Ramiro H. Gálvez

Mohamed Hafez, Malay Haldar, Mira Haldar, Hojae Han, Huimin Han, Shuguang Han, Maeda Hanafi, Jie Hao, Yongchang Hao, Zhuolin Hao, Mirazul Haque, Syed Mhamudul Hasan, Zia Hasan, Ashutosh Hathidara, Estrid He, Zirui He, Mohammad Hassan Heydari, Derrick Higgins, Wataru Hirota, Kasra Hosseini, Yubo Hou, Lingxiang Hu, Zhenyu Hu, Hanxian Huang, Jiahe Huang, Junbao Huang, Wei-Chieh Huang, Xinhao Huang, John S Hudzina, Madelon Hulsebos

Ignacio Iacobacci, Vladimir Iashin, Illarion Iov, Shotaro Ishihara, Sergei Ivanov, Arun Iyer

Mohammad Naser Sabet Jahromi, Eshaan Jain, Mayank Jain, Richa Jain, Srideepika Jayaraman, Seogyong Jeong, Feng Jiang, Hao Jiang, Jiawei Jiang, Jiye Jiang, Nan Jiang, Yankai Jiang, Zhida Jiang, Zhiwen Jiang, Zhuoxuan Jiang, Ishan Jindal, Yu Jinnai, Aviral Joshi, Warren Jouanneau, Prathyusha Jwalapuram, Huanglong ji

Muhammad Rafsan Kabir, Anubha Kabra, David Kaleko, Anup K. Kalia, Hidetaka Kamigaito, Madhav Kanda, Inho Kang, Jun Seok Kang, Vidhyakshaya Kannan, Diptesh Kanojia, Gautam Siddharth Kashyap, Preetham Reddy Kaukuntla, Simerjot Kaur, Pride Kavumba, Manish Kesarwani, Mansi Khemka, Dmitrii Khizbullin, Saksonita Khoeurn, Raad Khraishi, Geewook Kim, Honghoe Kim, HyeongSik Kim, Hyunsoo Kim, Seongwan Kim, Sunwoo Kim, G. Charbel N. Kindji, Tracy Holloway King, Elena Kochkina, Aakash Kolekar, Keshav Kolluru, Beibei Kong, Ana Kotarcic, George Kour, Rajasekar Krishnamurthy, Marek Kubis, Andrei Kucharavy, Shreyas Kulkarni, Abhishek Kumar, Amardeep Kumar, Ashwin Naresh Kumar, Dhruv Kumar, Gaurav Kumar, Pawan Kumar, Sanjeev Kumar, Tejaswini Kumar, Vineet Kumar, Gleb Kuzmin, Maksim Kuznetsov, Youngjun Kwak, Jihoon Kwon

Haoyang LI, Lujun LI, Yanis Labrak, Pratik Lahiri, Riwei Lai, Sarasi Lalithsena, Stefan Larson, Alexandra Lavrentovich, Duc V. Le, Hongseok Lee, Jongwuk Lee, Seunghan Lee, Sunkyung Lee, Shuo Lei, Yu Lei, Arun Balajiee Lekshmi Narayanan, Yves Lepage, Bei Li, Chong Li, Dingcheng Li, Fengyi Li, Mingchen Li, Qiang Li, Raymond Li, Ruiyu Li, Sha Li, Xingyu Li, Yanda Li, Yinghui Li, Zhenhao Li, Zhigen Li, Zhuang Li, Hao Liang, Veronica Liesaputra, Gilbert Lim, Woosang Lim, Nut Limsopatham, Antonie Lin, Ting-En Lin, Gui Ling, Yuan Ling, Angli Liu, Dongqi Liu, Hao Liu, Haowei Liu, Jiao Liu, Junzhuo Liu, Liqun Liu, Pengfei Liu, Rui Liu, Si-Yang Liu, Xingxian Liu, Xuye Liu, Yang Liu, Yang Liu, Ye Liu, Yimin Liu, Yonghao Liu, Zhaoxiang Liu, Yitao Long, Alejandra Lorenzo, Natalia V Loukachevitch, Guangming Lu, Wenlian Lu, Yuanjie Lyu, Wenwei li

Tianyi Ma, Xiangnan Ma, Zhiqiang Ma, Zhixin Ma, Abhinav Mahajan, Suraj Maharjan, Ayush Maheshwari, Rehana Mahfuz, Mohammed Makhoul, Valentin Malykh, Natalia Martinez, Lorenzo Masoero, Yuji Matsumoto, Hansa Meghwani, Mahnoosh Mehrabani, Yongsheng Mei, Kaiwen Men, Ge Meng, Samuel Mensah, Mohsen Mesgar, Chhaya Methani, Ruizhong Miao, Stuart E. Middleton, Hideya Mino, Katya Mirylenka, Lokesh Mishra, Vaishali Mishra, Hemant Misra, Sebastien Montella, Lori Moon, Kyosuke Morita, Yasmin Moslem, Lawrence S. Moss, Hamdy Mubarak, Koyel Mukherjee, Srijani Mukherjee, Sandeep Sricharan Mukku, Isaiah Onando Mulang', Matthew Mulholland, Prajit Reddy Muppidi

Shashi Bhushan Tyamagondlu Nagabhushan, Alwarappan Nakkiran, Sungjin Nam, Marcin Namysl, Diane Napolitano, Aravind Narayanan, Nilesch Nayan, Divya Neelagiri, Vera Neplenbroek, Günter Neumann, Chun Chet Ng, Huy V. Nguyen, Ping Nie, Yichen Nie, Prashanti Nilayam

Alexander O'Connor, Cennet Oguz, Oleg Okun, Arturo Oncevay, Hiroki Ouchi, Mete Ozay

Stefano Pacifico, Sayantan Pal, Manas Paldhe, Chester Palen-Michel, Abhinav Palia, Xueting Pan, Zijie Pan, Pierre-Henri Paris, JuneYoung Park, Abhay Dutt Paroha, Hima Patel, Kushal Patel, Meet Patel, Rohan G Patil, Vashishtha Patil, Anup Pattnaik, Matthias Paulik, Lucas Pavanelli, Vera Pavlova, Zhengqi Pei, Jianxiang Peng, Zhiyuan Peng, Ali Pesaranghader, Wang Piaohong, Siddharth Pillai, Qing Ping, Jakub Piskorski, Sanchita Porwal, Pradyot Prakash, Animesh Prasad, Hongxu Pu

Xuan Qi, Yiming Qian, Yiyue Qian, Jiayi Qiu, Xin Ying Qiu, Jiaming Qu, Yincen Qu, Elio Querze

Ahmed Y. Radwan, Ksheeraja Raghavan, Preethi Raghavan, Mizanur Rahman, Vyas Raina, Gopi Krishnan Rajbahadur, Akhil Ramachandran, Kiran Kumar Ramanna, Roshni Ramnani, Shihao Ran, Suraj Ranganath, Zixin Rao, Srijith Ravikumar, Traian Rebedea, Anton Repushko, Joy Rimchala, Brian Riordan, Anthony Rios, Lina M. Rojas-Barahona, Guoping Rong, Anup Roy, Chuanwei Ruan, Susanna Rücker

Wajdi Ben Saad, Alicia Sagae, Adithya Sagar, Horacio Saggion, Alsu Sagirova, Avishek Saha, Tanay Kumar Saha, Nutan Sahoo, Bhargobjyoti Saikia, Joy Prakash Sain, Andrey Sakhovskiy, Anoop Saladi, Md Sirajus Salekin, Luis Frentzen Salim, Avneesh Saluja, Madhumitha Saravanan, Yeahia Sarker, Prathusha K. Sarma, Minoru Sasaki, Andrey Savchenko, Andrea Scarinci, Frank Schilder, Phillip Schneider, Ulrich Schäfer, Ethan Selfridge, Srinivasan H. Sengamedu, Shubhashis Sengupta, Ronald Seoh, Sofia Serrano, Vinay Setty, Ebad Shabbir, Avijit Shah, Deven Santosh Shah, Fahad Shah, Jyoti Kunal Shah, Abdur R. Shahid, M. Ali Basha Shaik, Weiqiao Shan, Linfang Shang, Nitu Sharaff, Arpit Sharma, Umang Sharma, Shahin Shayandeh, Shivani Shekhar, Tianhao Shen, Xiang Shen, Yi Shen, Yu Shen, Tianyu Shi, Fatemeh Shiri, Egor Shvetsov, Shuzheng Si, Zohaib Hasan Siddiqui, Divyansh Singh, Harshvardhan Singh, Sanjeev Kumar Singh, Prateek Singhi, Gaurav Kumar Sinha, Sumit Soman, Hyun-Je Song, Xiaozhuang Song, Yewei Song, Nazib Sorathiya, Makesh Narsimhan Sreedhar, Arvind Krishna Sridhar, Varad Srivastava, Katherine Stasaski, Rebecca Steinert, Christopher M Stewart, Ian Stewart, Sebastian Stüker, Lixin Su, Lin Sun, Qi Sun, Yiding Sun, Yu Sun, Yuting Sun, Zhoujian Sun, Nataraj Agaram Sundar, Amit Surana, Munira Syed, Marek Šuppa

Anh Ta, Shingo Takamatsu, Sho Takase, Jayasimha Talur, Fei Tan, Jiwei Tan, Xuemei Tang, Kenil Yogeshkumar Tanna, Joel R. Tetreault, Ketan Thakkar, Aman Singh Thakur, Himanshu Thakur, Avinash Thangali, Sudarshan R. Thitte, Frank Thomas, Akshay Thorat, Onkar Thorat, Xing Tian, Matthew Toles, Yongqi Tong, Manabu Torii, Arthur Elwing Torres, Giuliano Tortoreto, Thanh Tran, Aashka Trivedi, Masaaki Tsuchida, Olga Tsymboi

Brian Ulicny, Mansi Uniyal

Venktesh V, Prasoon Varshney, Karthik Venkat Ramanan, Akshay Verma, Ishan Verma, Vidisha Vijay, Manuel Vilares Ferro, Andrei Vince, Maria Vlachou, Ngoc Phuoc An Vo, Elena Volodina, Xuan-Son Vu

Chaokun Wang, Chenglong Wang, Chunyang Wang, Chuyue Wang, Haoshen Wang, Jerry Wang, Jiale Wang, Jianzong Wang, Jihong Wang, Junfeng Wang, Lingxiang Wang, Peng Wang, Rongguang Wang, Ryan Wang, Wang Wang, Wenzhe Wang, Xinyu Wang, Yining Wang, Yu Wang, Yuxiang Wang, Yuying Wang, Zimu Wang, Zixuan Wang, Pankaj Wasnik, Changshuai Wei, Penghui Wei, Tengjin Weng, Julia White, Chenwei Wu, Fanyou Wu, Jikun Wu, Shenxi Wu, Taiqiang Wu, Tianxing Wu, Zekun Wu

Tianle Xia, Bo Xiao, Jie Xiao, Wei Xiao, Weihang Xiao, Xin-Yu Xiao, He Xie, Kaige Xie, Chuanji Xu, Jinan Xu, Mengyu Xu, Ming Xu, Xin Xu, Xin Xu, Zhiyu Xue

Ravindra Kumar Yadav, Rohan Kumar Yadav, Xiao Yan, Guanqun Yang, Yiming Yang, Yue Yang, Zhengzhe Yang, Junyi Yao, Xiao Yao, Yuhang Yao, Bingyang Ye, Dezhi Ye, Promod Yenigalla, Saurabh Yergattikar, Jinyeong Yim, Stella Xin Yin, Issei Yoshida, Cheng Yu, Chun-Nam Yu, Lei Yu, Qinkai Yu, Yipeng Yu, Zac Yu, Zewei Yu, Longfei Yun

Adeel Zafar, Mahdi Zakizadeh, Urchade Zaratiana, Kamyar Zeinalipour, Ziyin Zeng, Chi Zhang, Fuwei Zhang, Guilin Zhang, Haozhe Zhang, Huan Zhang, Jiahao Zhang, Jiangyun Zhang, Kai Zhang, Ke Zhang, Lei Zhang, Luxin Zhang, Mingxuan Zhang, Ningyu Zhang, Qianchi Zhang, Qingheng Zhang, Tianlin Zhang, Yifan Zhang, Yin Zhang, Yizhuo Zhang, Yue Zhang, Zhe Zhang, Ziyin Zhang, Cen Zhao, Fei Zhao, Haishu Zhao, Kai Zhao, Mengyue Zhao, Zijie Zhao, Maxim Zhelnin, Zihao Zheng, Zihao Zheng, Yunfei Zhong, Baohang Zhou, Dong Zhou, Dongzhuoran Zhou, Hongkuan Zhou, Peixian Zhou, Tianxin Zhou, Yifan Zhou, Youchao Zhou, Yu Zhou, Junfan Zhu, Junjie Zhu, Pengxu Zhu, Xi Zhu, Xiliang Zhu, Yuqicheng Zhu, Imed Zitouni, Hila Weisman Zohar, Ding Zou, Andrea Zugarini

Table of Contents

EMNLP 2026 Industry Track: Overview

Pepa Atanasova, Gholamreza Haffari and Minghan Wang 1

Beyond Offline A/B Testing: Context-Aware Agent Simulation for Recommender System Evaluation

Nicolas Bougie, Gian Maria Marconi, Xiaotong Ye and Narimawa Watanabe 5

LLP: LLM-Based Product Pricing in E-commerce

Hairu Wang, Sheng You, Qingheng Zhang, Xike Xie, Shuguang Han, Yuchen Wu, Fei Huang and Jufeng Chen 25

Prompts Don't Protect: Architectural Enforcement via MCP Proxy for LLM Tool Access Control

Rohith Uppala 43

Lunar-R1: Efficient Reasoning of Large Language Models for Lunar Exploration via Reinforcement Learning

Xin-Yu Xiao, Ye Tian, Zhixian He, Qilin Chen, Shiqi Wang and Qianchen Xia 50

Debate-to-Skill: Capability-Bound Process Supervision for Industrial Query-to-Agent Annotation

Shiyu Zhang, Leisheng Cheng and Huifu Li 70

Inventory-Grounded Policy-Level Optimization for Training-Free AI Search

Wei Zhou, Tiandeng Wu, Jiandong Ding, Zhufeng Fan and Yi Cao 79

.tmu: A Low-Entropy Tree-Structured Representation for LLM-Assisted Scientific Writing

Tianyou Liu, Ziqiang Li, Xurui Liu, Yu Wu and Yansong Li 100

How Much Should the Agent See? A Disclosure Axis for LLM Data-Analyst Agents

Athira M V, Aditya Challa, Jaya Tripathi, Saroj Mohanty, Santosh Kumar Choudhary, Sravan Danda, Ruchika Mehra Jain and Snehanishu Saha 124

Knsight: Self-Evolving Agentic System for Production Cloud Root Cause Analysis

Bo Lv, Hongge He and Xueqi Wang 145

GUIDE: Generative Unsupervised Chinese Query Correction via Phonetic and Visual Shared-ID Encoding

Lei Yang, Binbin Huang, Jiwei Tan, Xuhui Sui, Chang Tu, Yi Wang and Han Li 172

Pioneer Agent: Continual Improvement of Small Language Models in Production

Dhruv Atreja, Julia White, Nikhil Shivakumar Nayak, Kelton Zhang, Henrijs Princis, George Hurn-Maloney, Ash Lewis and Urchade Zaratiana 184

Cost-LLM: Efficient Multimodal Procurement-Cost Reasoning for E-Commerce Price Governance

Zhiyin Li, Fei Huang, Tengting Kong, Tang Jie and Shengjie Li 199

LocalSUG: City-Preference-Enhanced LLM for Query Suggestion in Local-Life Services

Jinwen Chen, Shiwen Zhang, Shuai Gong, Zheng Zhang, Yachao Zhao, Lingxiang Wang, Haibo Zhou, Wei Lin and Hainan Zhang 211

OmniThoughtVis: A Scalable Distillation Pipeline for Deployable Multimodal Reasoning Models

Yuanhao Yue, Chengyu Wang, Yuanjie Lyu, Lei Shen and Jun Huang 225

Complete or Abstain? Aligning SQL Completion with User Preferences

Xintong Hu, Defeng Xie, Yuxiang Wang, Chunyou Li, Haining Xie, Hao Liu, Haozhe Feng, Yang Li, Danqing Huang, Debiao He, Xiao Yan, Jiawei Jiang and Peng Chen 240

<i>Scaling Reinforcement Learning for Content Moderation with Large Language Models</i>	
Rui Liu, Yuchen Lu, Zhenyu Hou, Fangzhou Xiong, Xiaoyang Zhang, Changshu Jian, Zhicheng Zhu, Jiayuan Ma, Jacob Tao, Merve Unlu Menevse, Han Li, Chaitali Gupta, Xiaochang Peng, Suwen Lin, Chongyang Bai, Shike Mei, Hang Cui, Yang Qin, Shuo Tang, Jason Gaedtke, Arpit Mittal and Hamed Firooz	259
<i>Auditing and Repairing LLM-as-Judge Failures in a Production Text-to-SQL Pipeline</i>	
Haowei Liu, Hsin-Tai Wu and Yi Fang	275
<i>Plan or Retrieve First? A Scale Study of Tool Discovery Architectures</i>	
Mingfei Zhang, Amit Surana, Ziauddin Majid, Chenxi Wang and Michael Xu	288
<i>Toollery: Scaling LLM Agents to Thousands of Skills and Tools</i>	
Xiangxi Tian and Ran Guan	300
<i>RAFT: A Stateful Retrieval-Augmented Framework for Troubleshooting Agents</i>	
Mingxuan Zhang, Xiaowen Wang, Anupma Sharan, Zhengyi Chen, Chenyu Diana Zhang, Shanshan Yang and Chittibabu Pacharu	315
<i>PerceptUI: LLM Agents as Human-Aligned Synthetic Users for UI/UX Evaluation</i>	
Nicolas Bougie, Xiaotong Ye, Gian Maria Marconi and Narimawa Watanabe	329
<i>Multi-Stage Proxy Retrieval for Illiquid Bond Pricing</i>	
Arturo Oncevay, Joy Prakash Sain, Chen Ling, Simerjot Kaur, Charese Smiley and Xiaomo Liu	352
<i>Learning to Prompt: Improving Student Engagement with Adaptive LLM-based High-School Tutoring</i>	
Po-Chin Chang, Nicholas Hogan, Aske Laat and Michiel T. van der Meer	365
<i>WorkflowPerturb: Calibrated Stress Tests for Evaluating Multi-Agent Workflow Metrics</i>	
Madhav Kanda, Sharad Agarwal, Rodrigo Fonseca, Alok Gautam Kumbhare and Pedro Las-Casas	394
<i>TSAssistant: A Human-in-the-Loop Agentic Framework for Automated Target Safety Assessment</i>	
Xiaochen Zheng, Zhiwen Jiang, David Tokar, Yexiang Cheng, Alvaro Serra, Melanie Guerard, Klas Hatje and Tatyana Doktorova	412
<i>CityReal: Human-Aligned Urban Behavior and City Dynamics Simulation with Large-Scale LLM Agents</i>	
Nicolas Bougie, Xiaotong Ye and Narimawa Watanabe	435
<i>Parthenon Law: A Self-Evolving Legal-Agent Framework</i>	
Hejia Geng and Yuxuan Liu	456
<i>How LLMs Fail and Generalize in RTL Coding for Hardware Design?</i>	
Guan-Ting Liu, Chao-Han Huck Yang, Chenhui Deng, Brucek Khailany and Yu-Chiang Frank Wang	481
<i>SynthAVE: Scalable Synthetic Labeling for E-Commerce with LLM-Arena Validation</i>	
Andrea Scarinci, Virginia Negri, Brayan Impata, Suleiman A. Khan, Víctor Martínez and Marcello Federico	497
<i>Data Scout: Targeted Web Crawling for Domain-Specific Pretraining Corpora</i>	
Chirag Garg, Eelaaf Zahid, Farhan Ahmed, Jay Gala, Eric K Butler and Heiko Ludwig	519
<i>Interactor: Agentic RL oriented Iterative Creation for Ad Description Generation in Sponsored Search</i>	
Penghui Wei, Jiayu wu, Chao Ye, Zhi Guo, Shuanglong li and Liu Lin	532

<i>Mosaic: Automated Intent Description Generation for Multi-Label Classification</i>	
Ajay Chatterjee, Eric Crestan, Raunak Saha, Rotem Mordoch and Swarnangsu Acharyya . . .	545
<i>Tracing Target Answers in Poisoned Retrieval Corpora via Token Influence Attribution</i>	
Yan-Lun Chen, Pin-Yu Chen, Chia-Mu Yu, Ying-Dar Lin, Yu-Sung Wu and Wei-Bin Lee . . .	557
<i>FinStaR: Towards Financial Reasoning with Time Series Reasoning Models</i>	
Seunghan Lee, Jun Seo, Jaehoon Lee, Sungdong Yoo, Minjae Kim, Tae Yoon Lim, Dongwan Kang, Hwanil Choi, Soonyoung Lee and Wonbin Ahn	576
<i>Towards Fluid Consultations: Design and Real-World Evaluation of a Multi-Agent, Full-Duplex Health AI Assistant</i>	
Zhoujian Sun, Minghui Yang, Jian Wang, Jinjie Gu and Junwei Liu	610
<i>Quantization Breaks the Format: How Weight Compression Undermines Chat Format Safety in On-Device LLM</i>	
JuneYoung Park, Seongwan Kim and Jaeho Lee	626
<i>Beyond Semantic IDs: Encoding Business-Value Ranking into Document Identifiers for Generative Retrieval</i>	
Gui Ling, Zhihong Chen, Yu Li, Tong Xiong, Kunhai Lin, Kaixuan Zhang, Yuliang Yan, Dan Ou, Tanghaihong and Bo Zheng	641
<i>PULSAR: Pooled Unified Late-interaction Search And Retrieval for Enterprise Visual Document RAG</i>	
Ben Constable, Anup Roy, Vishal Sharma, Rishabh Upadhyay, Robin Mills and Aidan Millar	655
<i>Longitudinal-aware Evaluation of Retrieval-Augmented Generation over Evolving Credit Policy Documents</i>	
Kyosuke Morita, Jack Sanderson and Holly Hewitt	667
<i>ASSERT: A Measurement Pipeline for GenAI Audits</i>	
Riccardo Fogliato, Abhinav Palia, Xiawei Wang, Emily Sheng, Chad Atalla, Jean Garcia-Gathright, Nicholas Pangakis, Sharman Tan, Dan Vann, Hannah Washington, P. Alex Dow, Heba Elfardy, Hanna Wallach and Sandeep Atluri	681
<i>From Benchmarks to Production: A Text-to-SQL System for Complex Financial Data</i>	
Arijit Sehanobish, Bruno Gomes Coelho, Guillaume Michel, Sophia Zhi, Valerie Faucon-Morin and Kristen Howell	699
<i>SWE-Tester: Training Open-Source LLMs for Issue Reproduction in Real-World Repositories</i>	
Aditya Bharat Soni, Rajat Ghosh, Vaishnavi Bhargava, Valerie Chen and Debojyoti Dutta . . .	720
<i>Fidelity Is Not Enough: Dispatch-Level Instrumentation for Agentic Datasheet Extraction</i>	
Qing Ye and Meng-Hsuan Lin	730
<i>Quality Over Clicks: Iterative Reinforcement Learning for Early-Stage E-Commerce Query Suggestion</i>	
Qi Sun, Kejun Xiao, Huaipeng Zhao, Tao Luo and Xiaoyi Zeng	751
<i>From Simulation to Reality: Validating Automated Evaluation of Conversational Agents</i>	
Justyna Gromada, Ekaterina Isakova, Joanna Brudnik, Alicja Kasicka, Natalia Krawczyk, Bartosz Przybył, Lukasz Krajewski and Michał K. Szczerbak	763
<i>The Blocker Beasts and How to Tame Them: Schema-Driven Extraction in the Connective AI Layer of Network Operation</i>	
Madhava Gaikwad and Mohan Maddula	785

<i>MixEmb: Importance-Aware Quantization and Caching Co-Design for Industrial Decentralized Embedding Training</i>	
Zhida Jiang, Zhen Chen, Zhaolong Xing, Jiaxing Wang, Saixing Gao, Baopeng Yuan, Hao Wang, Yin Xu, Lin Liu, Dongyue Wang and Ke Zhang	803
<i>Session-Guided Pre-training for Query Understanding: Constraining Mask Prediction and Improving Query Fine-tunability</i>	
Sosuke Nishikawa and Naoki Yoshinaga	815
<i>Configurable Clinical Information Extraction with Agentic RAG: What Works, What Breaks, and Why</i>	
Osman Alperen Çinar-Koraş, Marie Bauer, Sameh Khattab, Merlin Engelke, Moon Kim, Stephan Settelmeier, Shigeyasu Sugawara, Fabian Freisleben, Felix Nensa and Jens Kleesiek	827
<i>FLASH-MAXSIM: IO-Aware Fused Kernels for Late-Interaction Retrieval</i>	
Roi Pony, Daniel Ezer, Adi Raz Goldfarb, Idan Friedman, Oshri Naparstek and Udi Barzelay	840
<i>Trains but Doesn't Learn: A Post-Training Delivery Benchmark for LLM Agents as Forward-Deployed Engineers</i>	
Weihang Ding and Junfei Zhan	868
<i>Loom: Weaving Diagnostic Strands into Free-Text Consensus via Embedding-Space Reweighting</i>	
Ron Begleiter, Katya Egert Berg, Gilad Saban and Gil Shabat	880
<i>SPAR: Experience-Driven Strategy Evolution in Multi-Session B2B Sales</i>	
Hangming Zhang, Yao Zhao, Haotian Kuang, C.L.Philip Chen, Tong Zhang and Tengyue Ma	892
<i>RAW: Refusal Aversion in the Wild, A Causal Measurement Method for Deployed LLMs</i>	
Pavel Kireyev	902
<i>What 1,114 Documents Reveal About Schema-Driven Extraction: Models, Pipelines, and the Real-vs-Synthetic Gap</i>	
Frank Thomas and Andrei Vince	928
<i>PriceBench: A Diagnostic Benchmark for Price, Quality, and Brand Preferences in LLM Booking Agents</i>	
Pavel Kireyev	941
<i>Towards Autopilot Labeling for Short Video Content Governance</i>	
Junwen Chen, Hongwei Wang, Zixuan Wang, Jingbing Wang, Mingchao Liu, Ong Sheng Siong, Xiaojiao Zang, Ju-Chin Chao, Ze Jin, Kaiwen Men, Yin Li, Jingwei Li, Zhuolin Hao, Hongyu Xiong and Yang Song	960
<i>A Practitioner's Guide to Systematic Prompt Evaluation for Information Extraction</i>	
Anna Steinberg Schulten, Malte Schierholz, Lisa Reichenbach, Frauke Kreuter and Alexander Fraser	975
<i>Closing the Operational Gap in Semantic Caching</i>	
Aditeya Baral, Radoslav Ralev, Iliya Zhechev, Srijith Rajamohan and Jen Agarwal	988
<i>EComAgentBench: Benchmarking Shopping Agents on Long-Horizon Tasks with Distributed Hidden Intent</i>	
Zeyao Du, Tong Li and Haibo Zhang	1012
<i>TripScore: Aligning LLMs for Real-World Travel Planning via Expert-Calibrated Reward</i>	
Yincen Qu, Huan Xiao, Feng Li, Gregory Li, Hui Zhou, Xiangying Dai and Huang Xuan	1028

<i>SkillChain: Closing the Loop on Skill Evolution for Image-Based E-Commerce AI Assistants</i>	
Yimin Hu, Mengtao Xu, Hao Guo, Yuheng Song, Xiaoyong Zhu and Bo Zheng	1055
<i>OpenDC-Infer-LC: Benchmarking Production-Valid Long-Context and RAG LLM Serving</i>	
Yue Yang, Maxwell Twelftree, Jared Collette, An-Chi He, Luke Borgnolo and David Lemphers	1070
<i>From Historical Queries to Actionable Rubrics: Category-Aware Rubric Generation for Local-Life AI Search</i>	
Jiale Mao, Fuwei Zhang, Dongbo Xi, Zhen Chen, Yifan Yang, Peng Yan, Zhao Zhang and Fuzhen Zhuang	1091
<i>PeopleSearchBench: Evaluating AI-Powered People Search Platforms with Criteria-Grounded Verification</i>	
Tianyu Shi, Wei Wang, Zequn Xie, Shuai Zhang, Boyang Xia, Chenyu Zeng, QI Zhang, Lynn Ai, YU Yaqi, Kaiming Zhang, Feiyue Tang, Zhenyu Yu and Lei Ding	1111
<i>Telco-GAIA: Bilingual Benchmark for Agents in Telecom Domain</i>	
Dmitrii Khizbullin, Zaid Alyafeai, Abdelrahman Eldesokey, Nourah AlSultan, Raghad Alshalan, Bernard Ghanem and David R. Pugh	1123
<i>Same Gains, Half the Cost: Selector Choice in Automatic Prompt Optimization for Financial Question Answering</i>	
Cyprian Bohojlo and Christopher M Stewart	1138
<i>EvidenT: Building Trustworthy Enterprise Assistants through Evidence Groundedness and Traceability</i>	
Anubha Kabra, Katie Jooyoung Kim, Colin Zhiwei Kou, Helene L Sajer, Yimei Fan, Radomir Cisar, Heather Greenhalgh and Gabriel Martinez Vidiri	1149
<i>RainbowPlus: Enhancing Adversarial Prompt Generation via Evolutionary Quality-Diversity Search</i>	
Quy-Anh Dang, Chris Ngo and Truong-Son Hy	1179
<i>Building Interpretable Feature Representations for Resume–Vacancy Matching by Distilling Production LLM Signals</i>	
Ilya Chekin, Vyacheslav Malyugin, Chirkov Vladimir and Mikhail Yurushkin	1209
<i>MortarBench: Evaluating Mortgage Loan Origination Agents</i>	
Matthew Toles, Yunan Lu, Manav Munjal, Bojun Liu, Yuanhao Deng, Stephanie Selig, Derek Rindner, Cheng Li and Zhou Yu	1222
<i>Retrieval-GRPO: Reinforcement Learning for Dense Retrieval in Taobao Search</i>	
Xingxian Liu, Dongshuai Li, Jiahui Wan, Tao Wen, Gui Ling, Fuyu Lv, Yuliang Yan, Dan Ou, Tanghaihong and Bo Zheng	1234
<i>ACE: A Self-Correcting Agentic Canvas Editor for Multi-Slide Presentation Automation</i>	
JooYoung Jang, Taegyeong Lee, Jihyeon Park and Nojun Kwak	1246
<i>Safe and Sufficient Caching: Validated Evidence Reuse for Production Retrieval-Augmented Generation</i>	
Akshay Verma, Siddharth Pillai, Sambit Das and Deepak Gupta	1272
<i>TaxCE : A Framework for Automated Taxonomy Construction and Evaluation at Scale</i>	
Sandeep Sricharan Mukku, Albert Aristotle Nanda and Rohit Pyati	1280
<i>Conversation Coach: A Voice-enabled AI System that Helps Practice Difficult Workplace Conversations</i>	
Fanyou Wu, Suraj Maharjan, Ainur Yessenalina, Dennis Xu Chen, Rahul Srivastava and Srinivasan H. Sengamedu	1302

<i>Deploying Programmatic Tool Calling for Large-Scale Agentic Systems</i>	
Nhan Nguyen, Hadi Habibzadeh and Ahmad Magdy Eid	1315
<i>Data-Driven Persona-Conditioned Agents for A/B Test Simulation</i>	
Ziyad Benomar, Weronika Łajewska, Leonardo Perelli and Saab Mansour	1323
<i>SCR-Eval: Evaluating Video Subtitle Translation with a Script-like Context Representation</i>	
Jongin Lee, Juyi Hwang and Sehyun Kim	1339
<i>APDMem: Agent-Controlled Progressive Disclosure for Query-Adaptive Long-Term Memory</i>	
Chin-Lun Fu, Anagha Kulkarni, Hong Ni and Behrouz Madahian	1361
<i>HyDRA: Hybrid Dynamic Routing Architecture for Heterogeneous LLM Pools</i>	
Aashna Garg, Siddharth Singha Roy, Shengyu Fu, Federico Brancasi, Jinu Jang and Giuseppe Cianci	1377
<i>Reasoning vs. Validation in LLM-based Pharmacological Parameter Extraction</i>	
Qinhao Wu and Xia Cui	1403
<i>MINCE: Shrinking LLM Evaluation Datasets via Few-Model Monte Carlo Calibration</i>	
Devleena Das, Rajeev Patwari, Vikram Kumar Bukka, Nithin Kumar Guggilla, Elliott Delaye and Ashish Sirasao	1418
<i>CHUNDER: Enabling Million-Token Qwen3 Serving on AMD ROCm</i>	
David Lemphers, Maxwell Twelftree, An-Chi He and Yue Yang	1429
<i>Beyond Query-Item Matching: Personalized Behavioral Relevance Modeling for E-commerce Search via Rubric-Guided Reinforcement Learning</i>	
Yang Liu, Chenxu Wang, Jianzhi Shao, Jun Ji, Xiaozhou Xu, Zhe Sun and Wei Hu	1444
<i>Cognitive Echoes: Deployable Conversational Risk Screening in Automotive Finance on Constrained LLMs</i>	
Zewei Yu and Yuting Sun	1456
<i>Techniques for Peak Memory Reduction for LoRA Fine-tuning of LLMs on Edge Devices</i>	
Hassan Dbouk, Matthias Reisser, Prathamesh Mandke, Likhita Arun Navali and Christos Louizos	1466
<i>Open Rubric System: Scaling Reinforcement Learning with Pairwise Adaptive Rubric</i>	
Wen Wang, Ruipeng Jia, Yunyi Yang, Yuxin Wu, Yongbo Gai, Siyuan Tao, Mengyu Zhou, Jianhe Lin, Xiaoxi Jiang and Guanjin Jiang	1501
<i>Retrieval-Conditional Parsing Score (RCPS): Choosing Document Parsers by Retrieval, Not by Appearance</i>	
Sang-Woo Son, Hyeong-seob Kim, Hyeonsang Kim, Hyun-woo Cho and Jinmo Kim	1531
<i>GraphReduce: Coverage-Preserving LLM Aggregation for E-commerce Review Insights</i>	
Hao Jiang, Xin Li, Yichi Zhang and Weisi Lin	1545
<i>Intent-Driven Generative Query Suggestion with Query-Level Credit Assignment</i>	
Xinpeng Liu, Lu Ma, Jiayi Qiao, Mengyu Zhou, Linglong Li, Xiaofeng Bian, Haonan Chen, Xiaoxi Jiang and Guanjin Jiang	1558
<i>CACHE-ED2: Compiling LLM Reasoning into Reusable Extraction Programs for Document Extraction at Scale</i>	
Sudhanshu Bhoi, Anurag Tripathi, Amir Raza and Mayank Jauhari	1570

<i>AgentLens: Trajectory Reviews for Coding Agent Evaluation</i>	
Andrey Podivilov, Vadim Lomshakov, Sergey Savin, Matvey Startsev, Roman Pozharskii, Maksim Parshin and Sergey Nikolenko	1588
<i>The Cost of Decomposition: Trade-Offs in Long-Context Aspect-Based Scientific Summarization</i>	
Pierre Achkar, Elena Senger, Tim Gollub, Martin Potthast and Yuri Campbell	1613
<i>Beyond Final Visibility: A Three-Stage Diagnostic Framework for Enterprise Generative Search Optimization</i>	
Charming	1637
<i>Querit-Reranker: Training Compact Multilingual Rerankers via Efficient Label-Free Distribution Adaptation</i>	
Yunfei Zhong, Jun Yang, Wei Huang, Yinqiong Cai, Haosheng Qian, Yixing Fan, Ruqing Zhang, Lixin Su, Daiting Shi and Jiafeng Guo	1648
<i>CemLit-IE: Schema-Constrained LLM Extraction for Traceable Cement Strength Databases</i>	
Sun Yanjie, Mengze Hong, Chen Jason Zhang and Haoyang LI	1663
<i>PRISM: Prompt-Refined In-Context System Modeling for Financial Retrieval</i>	
Chun Chet Ng, Jia Yu Lim, Zhen Hao Chu, Yixi Zhou and Low Wei Zeng	1674
<i>Rethinking Content Moderation via Severity-Aware Ranking</i>	
Yang Ou, Jielei Zhang, Yu Dai, Zian Yang, Run Yang, Xiang Li, Weihang Wang, Xueqian Zhao, Peiyi Li and Longwen Gao	1738
<i>FlowTrain: Flow-Based Decoupled Training for Industrial-Grade Vision-Language Models</i>	
Zhida Jiang, Zhaolong Xing, Pei Yang, Xiaolong Chen, Yuanhang Xiao, Chengzhi Huang, Xiyu Liu, Haopeng Liu, Qingyuan Sang, Lingfeng Zhou, Jiaying Wang, Zicheng Zhang, Wenzhe Wang, Liuxinyu, Yan Li, Zhen Chen and Ke Zhang	1754
<i>Predefined-Tool Execution for Reliable LLM Agents in Semiconductor Lot Analysis</i>	
Nirupama Vijayaraghavan Malolan and Duc V. Le	1765
<i>TeleTables: A Benchmark for Evaluating LLM Understanding of Tables in Technical Standards</i>	
Anas Ezzakri, Nicola Piovesan, Mohamed Sana, Antonio De Domenico, Fadhel Ayed and Haozhe Zhang	1788
<i>FINSTRUCT: LLM-Based Logical Structure Recovery for Financial Filings</i>	
Mazen Mohamed Nasr Mohamed Ahmed, Yunfei Long, Michael Kampouridis, Rachel Tattersall and Kimberley Clague	1804
<i>SkillHone: A Harness for Continual Agent Skill Evolution Through Persistent Decision History</i>	
Zhiwei Li and Yong Hu	1819
<i>GRAM-V2: Generative Reasoning and Ranking Alignment for Large-Scale E-commerce Search</i>	
Chunyan Yuan, Ming Pang, Yiming Deng, Xiaoyu He, Zheng Fang, Donghao Xie, Fanyi Qu, Changping Peng and Ching Law	1830
<i>D2GS: A Data-Centric Framework for LLM-based Creative Game Script Generation</i>	
Jiali Wang, Junbao Huang and Minghao Chen	1840
<i>DiagFlowBench: Evaluating How Language Models Handle Off-Procedure Inputs in Grounded Diagnostic Dialogue</i>	
Guillermo Gil de Avallé, Laura Maruster, Shaina Raza and Christos Emmanouilidis	1854

<i>Style as a Confound: False Positives in AI Detection of Non-Native Academic Writing</i>	
Hyeonchu Park, Gahye Jeong and Bugeun Kim	1872
<i>When Not to Think: Mitigating the Overthinking Effect in Content Moderation</i>	
Yu Dai, Jielei Zhang, Yang Ou, Zian Yang, Run Yang, Xiang Li, Qingyao Zhuang, Peiyi Li and Longwen Gao	1888
<i>EAGER: Enrich-and-Align Generative Query Recommendation from Clicked Items in E-commerce Search</i>	
Shuwei Yuan, Mingqian Ding, Luxin Liu, Rong Xiao and Xiaoyi Zeng	1900
<i>Perceptually Better, Semantically Worse: Measuring Speech Enhancement Impact on LLM-Based Voice Systems</i>	
Randy Frans Fela and Pejman Mowlae	1913
<i>AuDQ: Auditable Data Quality through Agentic Rule Discovery</i>	
Jacopo Biggiogera, Nicolas Colombo, Krati Saxena, Bhaskarjit Sarmah and Stefano Pasquali	1925
<i>Voice of Australian-AI: Learning Australian Assistant Voice Without Caricature</i>	
Yue Yang, Maxwell Twelftree, Luke Borgnolo, David Lemphers and Amanda Baughan	1944
<i>Less Is More: Graph-free Multimodal RAG via Multi-signal Late Fusion</i>	
Tithi Rakshit, Hongkuan Zhou, Lavdim Halilaj and Yuqicheng Zhu	1962
<i>Lightweight Temporal Attention for Video Retrieval over Frozen Image and Text Encoders</i>	
Vidit Khazanchi, Umang Sharma and Shabhareesh Komirishetty	1976
<i>From Production Traffic to Post-Training: Building a Self-Hosted LLM That Covers the Corporate Request Mix</i>	
Olga Tsymboi, Dmitrii Stoianov, Ramil Latypov, Danil Taranets, Daniil Dryabin, Mikhail Gashkov, Viktor Zelenkovskiy, Aleksandr Fida, Gleb Alektorov, Nikita Gulyakov, Arthur Babkin, Aleksandr Medvedev, Pavel Gein and Anatolii Potapov	1995
<i>Closing Cost-Quality Gap in Document VLMs: Difficulty-Aware Data Curation and Quality-Adjusted Deployment Economics</i>	
Maksim Evdokimov, Matvey Ivanov, Dmitrii Tsiupin, Olga Tsymboi, Anatolii Potapov and Aleksandr Ivanov	2024
<i>Business Compromise Detection with Agentic AI and LLM-driven Knowledge Discovery</i>	
Diego Palma, Kyu Bin Kim, Zhen Han, Allbright Dsouza and Zhiyuan Liu	2045
<i>Benchmarking AI Agents for Hardware Design Automation via MCP Tool Calling</i>	
Leonardo Liparulo and Francesco Pierri	2058
<i>From Synthetic to Localized: Benchmarking Multilingual Intent Classification in Logistics Customer Service</i>	
Haoyu He, Jinyu Zhuang, Haoran Chu, Shuhang yu, Hao Wang and Han Kunpeng	2078
<i>Adaptation, Not Translation: An Architecture and Evaluation Framework for Long-Form Cross-Cultural Story Localization</i>	
Taaha Kazi, Aditya Sanjiv Kanade, Ivory Yang, Mohammad Saifullah, Vasu Sharma, Bhavesh Pant, Kunal Bhatia, Abhinav Prakash, Abdur Rahman, Varun Agarwal, Sarjita Jain, Viveca Pavate and Nishita Maithul	2093
<i>Beyond Accuracy: Robustness, Cost, and Governance Trade-offs for Vision-Language Models in Templated Document Extraction</i>	
Kushal Patel, Pushkal Shrivastava, Mackenzie Lees, Qirui Lu, Bhargobjyoti Saikia, Liying Li and Junlin Jiang	2115

<i>Schema-Hierarchical Soft Prompts for Structured Information Extraction</i>	
Soumyadeep Dey, Srikanth Tenneti, Bamdev Mishra, Shobhit Niranjana and Mayur Datar . . .	2127
<i>From Confusion to Clarity: Confusion-Aware Retrieval and Knowledge Injection for Text Classification</i>	
Manish Gupta, Chaitanya Giri and Jayasimha Talur	2149
<i>RLMA: Reinforcement Learning with PolyMetric Alignment</i>	
Jia Xu	2165
<i>Edit4Agency: Efficiently Enhancing Agent Sub-Capabilities of Small Language Models via Model Editing</i>	
Jianxiang Peng, Yue Chen, Zhigen Li, Yanmeng Wang, Ning Cheng, Shaojun Wang, Yuqi Ren and Deyi Xiong	2183
<i>SAFARI: An Industrial Benchmark for LLM-Assisted Hazard Analysis and Risk Assessment</i>	
Chenxi Wu, Zimu Wang, Haiyang Zhang, Wei Wang and Zhijie Xu	2197
<i>SCOPE: Stratum-Coordinated Optimization Pipeline for Efficient Agent Context</i>	
Pushpendu Ghosh, Aatir Zaki, Rudraraju Sacin Varma, Stanley Kanagaraj and Anish Nediyan-chath	2207
<i>Human-Anchored Factuality Evaluation with Strategic Annotation</i>	
Yu Wang, Craig Erickson and Kevin Small	2218
<i>Attention-Weighted Retrieval-Augmented Data Engineering for ADR Severity Classification</i>	
David Guo, Nishant Vishwamitra and Kim-Kwang Raymond Choo	2234
<i>Enrich-Retrieve-Rank: Scaling Capability Discovery Beyond In-Context Routing</i>	
Nazib Sorathiya, Daniel Zhang and Bardiya Akhbari	2244
<i>GLS Chatbot: A Governed Multi-Agent Conversational System for Domain-Specific Customer Service</i>	
Bernat Coma-Puig, Guillermo Andrés Galván Hernández, Victor Mas Montserrat, Edgar Sarria Tenes, David Niñoles Asin and Sonia Alonso Montoya	2255
<i>REPO2ROCM: An Agentic System for Reproducing Research Code on AMD ROCm GPUs</i>	
Gaurav Sarkar, Umang Pandey, Pratik Prabhanjan Brahma, Zicheng Liu and Emad Barsoum	2265
<i>TRACE: Traceable Root-cause Analysis with Calibrated Evidence-grounded agents</i>	
Aniket Joshi, Cyrus Andre DSouza, Sejal Jain and Manjunath Channappagoudar	2279
<i>Mem-LM: An LM-style Model for Reinforced Memory Selection</i>	
Yanhan Huang, Yihao Zhong, Wenwen Ouyang, Xing Tian, Zining Yin and Yiyi Lu	2290
<i>From Metrics to Natural Dialogue: French Full-Duplex Benchmark for Spoken Dialogue Models</i>	
Hamid Soltani and Gilles Boulianne	2302
<i>BIRD-Persona: A Persona-Aware Benchmark for SQL and pandas Analytics Code Generation</i>	
Davin Hill, Pawan Chorasaya, Bhawani Shankar Leelar, Oren Avram, Robert E. Tillman and Tamer Soliman	2311
<i>EGT-KG: Evidence-Grounded Typed KG Retrieval for Practical Scientific QA with Small Language Models</i>	
Muran Yu, Jiechao Gao, Yuandong Pan, Barney Haoyun Miao, Andrew Carter Lesh, Kincho Law, Jie Wang and Michael Lepech	2333

<i>Small Model, Big Retrieval: Unified Heterogeneous Retrieval via Multi-Label-Aware Contrastive Fine-Tuning</i>	
Shaobai Jiang, Dantong Liu, Federico Gonzalez, Pedro Herrero-Vidal, You-Lin Chen, Xiaoyu Chu and Prithviraj Sen	2355
<i>An Empirical Study of VLM Pipelines for Long-Document QA</i>	
Kenan E. Ak, Jay Mohta, Gwang Gook Lee, Yan Xu and Dimitrios Dimitriadis	2369
<i>StreamCorrect: Self-Supervised Error Correction for Real-Time ASR with Offline Recognizers</i>	
Tung D. Vu, Zeying Xie, Weiwei Zhao, Victor Junqiu Wei, Di Jiang and Yi R. Fung	2391
<i>Better Together: Complementary Query Rewriting Under a Strong RAG Baseline</i>	
Sara Shanian, Xiaoqin Yi, Pavlo Ruban and Kurt MacDonald	2404
<i>Script Gap: Evaluating LLM Triage on Indian Languages in Native vs Romanized Scripts in a Real-World Setting</i>	
Manurag Khullar, Utkarsh Desai, Poorva Malviya, Aman Dalmia and Zheyuan Ryan Shi . . .	2414
<i>What Signal Does Preemption Need? Benchmarking Preemption-Target Selection for SLO-Aware Reasoning LLM Serving</i>	
Qi Yang Huang, Alay Dilipbhai Shah, Pragaash Ponnusamy, Hoang Anh Duy Le, Thomas A. Joseph, Adam Seligman and Zhaozhuo Xu	2430
<i>When LLM Plans Fail: Solver-Verified Remediation Under Operational Constraints</i>	
Priyaranjan Pattnayak	2453
<i>ForensicBench: Evaluating Agentic LLMs on Journal-Entry Fraud Detection</i>	
Guy Stéphane Waffo Dzuyo, Gaël Guibon, Christophe Cerisara and Luis Belmar Letelier . .	2473
<i>Cheaper Than Frames: An OCR+Keyframe Index as a Perception Layer for GUI-Video Question Answering</i>	
Rushikesh Hiray, Shubham Srivastava, Azizur Rahman, Laraib Hasan, Mohammad Hijjaw and Harshit Taneja	2489
<i>Data Intelligence Agents: Interpreting, Modeling, and Querying Enterprise Data via Autonomous Coding Agents</i>	
Anoushka Vyas, Aarushi Dhanuka, Sina Khoshfetrat Pakazad and Henrik Ohlsson	2500
<i>Re:CAP - Auditing Retrieval Coverage in Production RAG Pipelines</i>	
Aviral Joshi, Hanoz Bhathena, Max Nelson and Saket Sharma	2532
<i>FinDialogLens: Event Extraction over Multi-Party Dialogue for Missed-Trade Identification in Financial Chatrooms</i>	
Chin-Lun Fu, Hong Ni and Behrouz Madahian	2566
<i>A Hybrid Framework for Precedent Reference Extraction</i>	
Arthur Elwing Torres, Johny Moreira, Daniela Vianna, Marina Ramallete Masid, Celio Larcher, Edleno Silva de Moura and Altigran da Silva	2582
<i>From Monolithic Blending to Agentic Orchestration: Dynamic Response for Conversational Assistants at Scale</i>	
Cen Zhao, Peng Wang, Chuan Shi, Yufeng Zhang, Ying Lyu, Wanmeng Ren, Robert Xue, Claire Cheng and Yashar Mehdad	2591
<i>Tool-Making and Self-Evolving LLM Agents in Low-Latency Systems</i>	
Kalle Kujanpää, Ning Liu, Shahnawaz Alam, Yeshwanth Reddy Sura, Tianyu Yang, Kristina Lisa Klinkner and Shervin Malmasi	2608

PLSQLBench: Benchmarking LLM Systems for Executable Procedural Database Programming

Marianne Menglin Liu, Leonid Boytsov, Daniel W. Peterson, Pramuditha Perera, Rongguang Wang, Sai Ashish Somayajula, Syed Hamza Rafique, Rohit Saini, Shubham Pathak, Sujeeth Bharadwaj, Tao Sheng, Graham Horwood, Ankan Bansal, Fahad Shah, Sujith Ravi and Dan Roth 2625

Frustratingly Easy Agentic RAG via Portable Skills: A Case Study in Customer Support

Istok Bako, Bonan Min, Qihong Shao, Claire Cheng and Yashar Mehdad 2647

FlashTrie: A GPU-Accelerated Constrained Beam Search for Generative Retrieval

Dakshitha B Anandakumar, Anurag Mukkara, Wenxiang Hu, Akash Kumar Mohankumar, Jiusheng Chen, Qiang Lou, Jian Jiao and Ting Ye 2662

WARP: Wasserstein-Aligned RAG for Population Opinions

Aman Singh Thakur, Aditya Agrawal, Alwarappan Nakkiran and Alex Karlsson 2684

Eluna: An Agentic LLM System for Automating Warehouse Operations with Reasoning and Task Execution

Ning Liu, P Aditya Sreekar, Kalle Kujanpää, Zhaoxuan Zhu, Kaiwen Liu, Chuanneng Sun, Jorge Marchena Menendez, Matthew J Bales, Tianyu Yang, Shah Nawaz Alam, Rose Yu, Baoyuan Liu, Kristina Lisa Klinkner and Shervin Malmasi 2705

Leveraging Turn-taking Dynamics for Intent Recognition in Multi-party Conversations

Galo Castillo-López, Alexis Lombard, Gaël de Chalendar and Nasredine Semmar 2717

UltraQuant: 4-bit KV Caching for Context-Heavy Agents

Inesh Chakrabarti, David Limpus, Aditi Ghai Rana, Bowen Bao, Spandan Tiwari, Thiago Crepaldi and Ashish Sirasao 2732

Judge, Don't Write: Scaling Golden Data Generation via Psychometric Standard-Setting

Scott Novotney, Youbing Yin and Erin Babinsky 2751

A Pinch of SFT, A Dash of RL: When Reinforcement Learning Helps Long-Horizon Advertising Agents

Aakash Kolekar, Sahika Genc, Bunyamin Sisman, Shahriar Shariat, Shree Vandana Kachroo, Avishek Saha, Qianli Wu, Ari Singer and Benoit Dumoulin 2762

SEGRA: A Structured Experience Guided Reasoning Agent for Property Graph Question Answering

Saiyue Lyu, Mariam Dundua, Vishaal Kapoor, Sarthak Ahuja, Neda Kordjazi, Evren Yortucboyly, Harsh Amin and Rebecca Steinert 2779

From Benchmarks to Real-World Use: Evaluating the Robustness of Vision-Language Models for Chart Understanding

Awal Ahmed Fime, Mohammed Saidul Islam, Md Tahmid Rahman Laskar, Md Zarif Hossain, Ridwan Mahbub, Mir Tafseer Nayeem, Mizanur Rahman, Ahmed Imteaj and Enamul Hoque Prince 2794

The Differential Reasoning Router: Operationalizing Cost-Aware LLM Annotation in E-commerce

Cheng Lyu, Jingyue Zhang, Vinny DeGenova, Mengwei Li and Yuanli Pei 2815

RCEM: Robust Conversational Search Embedder in Distributional Shift

Kilho Son, Bo-June Paul Hsu, Cha Zhang and Dinei Florencio 2830

PROOF-Gen: From Optimized Data to Better Distillation

Anh Ta, Junjie Zhu and Shahin Shayandeh 2844

Peerify: Benchmarking Peer-Review Claim Verification

Alireza Daghighfarsoodeh, Sajad Ebrahimi, Ali Ghorbanpour, Soroush Sadeghian, Radin Cheraghi, Negar Arabzadeh and Ebrahim Bagheri 2865

<i>Extending FunctionGemma for Practical On-Device Mobile Function Calling</i>	
Ali Rezagholizadeh and Soheila Samiee	2889
<i>CONFRACT: Confidence-Guided Retrieval-Augmented Generation</i>	
Yin Huang, Yifan Ethan Xu, Kai Sun, Vera Yan, Alicia Yi Sun, Haidar Khan, Jimmy Nguyen, Jingxiang Chen, Mohammad Kachuee, Zhaojiang Lin, Yue Liu, Aaron Colak, Anuj Kumar, Wen-tau Yih and Xin Luna Dong	2905
<i>When Synthetic Data Hurts: On Catastrophic Forgetting in Skill Retrieval for LLM Agents</i>	
Syed Shariyar Murtaza, Yifan Nie, Utkarsh Soni, Eugene Wen and Arvid Frydenlund	2924
<i>RECOVER: Robust Entity Correction via agentic Orchestration of hypothesis Variants for Evidence-based Recovery</i>	
Abhishek Kumar and Aashraya Sachdeva	2939
<i>From Text to Alpha: Can LLMs Track Evolving Signals in Corporate Disclosures?</i>	
Chanyeol Choi, Yoon Kim, Yu Yu, Young Cha, Zach Golkhou, Igor Halperin, Georgios Papaioannou, Minkyu Kim, Zhangyang Wang, Jihoon Kwon, Minjae Kim, Alejandro Lopez-Lira and Yongjae Lee	2952
<i>ChemCLIR-Bench: Benchmarking Cross-Lingual Information Retrieval in Multilingual Chemical Patents</i>	
Mahdi Astaraki, Mohammad Khodadad, Reza Namazi, Mohammad Arshi Saloot, Amir Reza Behzad Moghadam, Hamidreza Mahyar and Soheila Samiee	2962
<i>LivingArena: Do LLMs Know What Other LLMs Don't? Peer-Probing as Scalable Evaluation</i>	
Xingyu Chen, Rui Wang, Zhaopeng Tu and Liefeng Bo	2984
<i>Learning What Not to Forget: Long-Horizon Agent Memory from a Few Kilobytes of Learning</i>	
Nusrat Jahan Lia and Aritra Mazumder	2999
<i>Displacement Is Not Direction: Evaluating Fidelity Metrics for Quantized LLM Deployment</i>	
Milos Nikolic, Ali Hadi Zadeh, Enrique Torres Sanchez and Andreas Moshovos	3025
<i>Scaling Enterprise Cloud Agents to Thousands of APIs: An Empirical Study of Agent Architecture and MCP Tool Boundaries</i>	
Sandip Ghoshal, Khalil Mrini, Maram Assi, Yadan Fan, Jyotika Singh, Miguel Ballesteros, Weiyi Sun, Anshul Mittal, Yassine Benajiba, Sujeeth Bharadwaj, Sujith Ravi and Dan Roth	3040
<i>When Does User-Scoped Glossary Injection Recover ASR-Corrupted Domain Terms in Meeting Summarization? Separating Term Recovery from Summary Utility</i>	
Gihwan Kim, Hyoungh Jun Lim, Yoonseo Chung, Heeyong Park, ChangHyeon Jeong, Seunghee Ma, Joohye Park and Junghwan Kim	3056
<i>PrismAlign: Prior-Steered Multi-View VLM Alignment for Hallucination-Robust Table OCR</i>	
Guangyi Liu, Qianjun Huang and Boyu Hou	3069
<i>Decoupling Search from Reasoning: A Vendor-Agnostic Grounding Architecture for LLM Agents</i>	
Emmanuel Aboah Boateng, Kyle MacDonald, Amardeep Kumar, Siddharth Kodwani and Sudeep Das	3079
<i>DI-Bench: Systematically Generating In-Domain Data Intelligence Benchmarks for Enterprise Agents</i>	
Jiangyun Zhang, Kristen Surrao, Tor Nitayanont, Yupei Zhang, Roopali Singh, Zhiyu Chen, Julia Huang, Zhou Tang, Shayan Ali Akbar, Omar Alonso, Erwin Cornejo, Yuan Li and Yi Zhang	3095
<i>Vision is Not Enough: Enhancing Bounding Box Grounding for Document Key Information Extraction with OCR Anchors</i>	
Zhaochen Li, Rathika Suresh Kumar, Tao Bai, Zhijie Zhang, Hongxin Shao and Daniel Dahlmeier	3114

<i>From Diverse Outputs to Structured Set: Semantic Planning Enables Parallel Advertising Text Generation</i>	
Penghui Wei and Jiayu wu	3127
<i>Point-in-Time Financial RAG with Frozen LLM Readers and Market-Feedback Adaptive Retrieval</i>	
Zijie Zhao and Roy E. Welsch	3137
<i>When to Defer: Scope-Aware Cascades for Structured Attribute Generation in E-commerce Product Catalogs</i>	
Fatemeh Mansoori, Andrea Scarinci, Aditya Aggarwal, Ashwin Chandramouli and Suleiman A. Khan	3149
<i>Scaling RLHF Annotation: A Hybrid Human-AI Collaboration Framework</i>	
Khanh Nguyen, Shi Chen, Suseella Panguluri, Jesse Oliver, Salika Dave, Pritam Gundecha and Olivier Delalleau	3161
<i>AdaMem: Learning What to Remember with Adaptive Memory Policies for Personalized Agents</i>	
Xingyu Chen, Rui Wang, Zhaopeng Tu and Liefeng Bo	3186
<i>GAUGE: When Not to Trust LLM-as-a-Judge in User-Simulated Evaluation of Task-Oriented Agents</i>	
Umesh Bodhwani, Thanh Tran and Kai Wei	3197
<i>OmniQUM: Multilingual Query-to-Location Understanding Model for Local Search</i>	
Avijit Shah, Srini Dwarakanathan, Yash Kamdar, Danielle Class and Alessandro Moschitti ..	3217
<i>Positional task conditioning for scalable defect detection across product families in large product catalogs</i>	
Soham Satyadharma, Gabriel Roccabruna and Suleiman A. Khan	3226
<i>When Evaluators Cry Wolf: Lessons from Production LLM-as-Judge Evaluation in Educational AI</i>	
Chris Rohlf, Rodrigo Vergara Bosse, Priscilla Hopper, Keanon O’Keefe and Patrick J Russell	3240
<i>Litmus: Zero-Label, Code-Driven Metric Specification for Evaluating AI Systems</i>	
Prajwal Gupta, Prasang Gupta, Vishal Bhutani, Apoorva Sharma, Sumanth Chundru, Waqar Sarguroh and Kevin Paul	3256
<i>VoiceAgentGuard: Contract-Gated Tool Release for Production Voice Agents</i>	
Emmanuelle Bourigault and Pauline Bourigault	3281
<i>Iterate Until Retrieved: Factual Nugget Optimization for Discoverable Continual Corrections in Agentic RAG</i>	
Moshe Hazoom, Gal Patel, Alon Talmor and Tom Hope	3297
<i>SIRF: A Spec-Internalized Risk Foundation Model for Industrial Content Risk Control</i>	
Suwan Wu, Yumeng Lin, Pengcheng Yuan and Xiaolong Jiang	3308
<i>Not All Tokens Earn the Reward: Contrastive Credit Assignment for Production Detection</i>	
Tejas Alisyam, Prateek Sircar and Deepak Gupta	3322
<i>Awakening Dormant Users: Modeling Intent Evolution via LLM-based Functional Role Reasoning for Generative Recommendation</i>	
Huishi Luo, Shuokai Li, Hanchen Yang, Zhongbo Sun, Haojie Ding, Boheng Zhang, Zijia Cai, Renliang Qian, Fan Yang, Tingting Gao, Chenyi Lei, Wenwu Ou and Fuzhen Zhuang	3334
<i>Do Agents Help? Agentic Document Information Extraction with Enterprise Database Grounding</i>	
Zhijie Zhang, Hongxin Shao, Mattis Lennart Geiger, Tao Bai and Daniel Dahlmeier	3357

<i>When to Trust: Margin-Calibrated Routing Between LLM-as-a-Judge and Human Experts</i>	
Xinran Zhang, Ping Nie, Michelle Lok-ye Chan, Joshua L. Mason, Arash Sadati and Tate Sterling Avery	3376
<i>DynamicMCPBench: A Trace-Grounded, Effect-Scored Benchmark for LLM Agents over Live MCP Servers</i>	
Jerzy Kamiński, Ilya Galyukshev, Artem Kuznetsov, Sergey Chuprin, Kirill Redko, Aidar Shumbalov and Anna Kalyuzhnaya	3391
<i>GraphBit: A Graph-based Agentic Framework for Non-Linear Agent Orchestration</i>	
Yeahia Sarker, Md Rahmat Ullah, Musa Molla and Shafiq Joty	3412
<i>REDACT: A Systematically Controlled Multilingual Benchmark for Personal Information Detection</i>	
Guneesh Vats, Anubha Agrawal, Shikha Singhal, Ajita Dash, Praisson Selvaraj, Vidhan Jhavar, Ranga Prasad Chenna and Bharadwaj Y M G	3426
<i>CrisisLink: Closing the Loop between Underspecified Crisis Posts and Humanitarian Coordination</i>	
Yuchen Zhang, Yuze Gao, Bin Chen, Jian Su and Farah Benamara	3441
<i>How to Speculate about Uncertainty in Agentic Coding? A Draft-Model Gate Method</i>	
Konstantin Grotov and Valentin Malykh	3455
<i>Beyond Dense Retrieval: A Structured-Semantic Approach for Conversation Analytics</i>	
Abinash Kanagarajan, Chetan Aggarwal and Anjali Jain	3466
<i>ProChunker: Self-Optimizing Procedural Chunking for Agent-Ready Workflow Graphs</i>	
Soumajit Roy, Sayanta Adhikari, Vinayak S Puranik and Prakash Mandayam Comar	3490
<i>RACA-PODS: Reasoning-Informed First-Token Prediction for Query-Document Relevance in WeChat Search</i>	
Yupeng Hu, Dongliang Liao, Changxing Ding, Mengchuan Qiu, Huasheng Liang, Qiang Yan and Junhong Wang	3520
<i>ONTOFORM: Ontology-Grounded Semantic Data Generation for Enterprise Document Understanding</i>	
Nilakshan Kunanantaseelan, Benjamin Kane, Siddharth Jain, Mahsa Ghorbanali, Fatemeh Shiri, Raj Tumuluri and Gholamreza Haffari	3535
<i>FedAnchor: Differentially Private Federated LoRA via Weight Anchoring</i>	
Jie Xu, Haaris Mehmood, Rogier van Dalen, Karthikeyan Saravanan and Mete Ozay	3544
<i>LLM Post-Training as Brownfield Maintenance: An Industrial Perspective on Dataware Engineering</i>	
Gopi Krishnan Rajbahadur, Amir Ebrahimi, Boyuan Chen and Ahmed E. Hassan	3555
<i>Predicate Importance Estimation and Decoupled Rationale-Score Distillation for Entity Alignment</i>	
Keunha Kim, Yoonjin Jang, Hyeon-gu Lee, Sihyung Kim and Youngjoong Ko	3570
<i>Adapt the Words, Freeze the Model: LLM Prompt Evolution for Domain-Shift-Robust Segmentation</i>	
Olga Volkova, Nikita Glazkov, Pchelin Konstantin, Iaroslav Radionovich Bepalov, Evgeny Evgenievich Baraboshkin, Denis Orlov and Dmitry V. Dylov	3588
<i>AutoGraphQL: A Bring-Your-Own-Schema Framework for Adapting Open-Weight LLMs to GraphQL APIs</i>	
Keya Battu, Parth Mehta, Harikrishnan Balagopal and Manish Kesarwani	3602
<i>Relevance is not enough: A Communication-Oriented Retrieval System for Consequential Scientific Question Answering</i>	
Avina Nakarmi, Naga Datha Saikiran Battula, Anthony Diaz and Aritra Dasgupta	3614

<i>LDU-Bench: Multimodal LLM Evaluation for Lithography Defect Understanding under Layout-Varying Circuit Backgrounds</i>	
Huanglong ji, Botong Zhao, Shujing Lyu and Yue Lu	3626
<i>MERIT: Mitigating Exposure Bias in Generative XMC for User-Interest Propensity Modeling</i>	
Abhinav Mahajan, Arindam Sarkar and Prakash Mandayam Comar	3638
<i>EnterpriseClawBench: Building Enterprise Agent Benchmarks from Real Workplace Sessions</i>	
Jincheng Zhong, Weizhi Wang, Che Jiang, Kai Tian, Zhenzhao Yuan, Junlin Yang, Dianqiao Lei and Kaiyan Zhang	3657
<i>Med-GRADE: Medical Grading and Rubric-based Assessment of Doctor-Patient Encounters</i>	
Yidong Jiang, Chao Ding, Renjie lu, Rongzhao Zhang, Yankai Jiang, Tianbin Li, Junzhi Ning, Chengze Li, Haoyu Zhang, Jinru Ding and Pengcheng Chen	3673
<i>SAGE: Stochastic Prompt Optimization via Agent-Guided Exploration</i>	
Ziyi Zhu, Luka Smyth, Saki Shinoda and Jinghong Chen	3695
<i>Multi-View Structural Interpretability of Agentic Reasoning Traces</i>	
Ritika Hooda, Sravanthi Machcha, Jackson MacDonald, Lavanika Srinivasaraghavan, Atharva Nijasure, Rui Wang, Lei Shi, Junda Wu and Tong Yu	3712
<i>AraDynFact: Dynamic Evaluation of Factual Knowledge in Arabic</i>	
Ignacio Iacobacci, Farooq AL-Tam, Zhaozhi Qian and Muhammad AL-Qurishi	3741
<i>Efficient Multi-Tenant LoRA Serving for Encoder Models via Late-Fusion Batched Matrix Multiplication</i>	
Ashutosh Rupoliya Dwivedi, Keshav Shrikant and Jyotsna Rupaliya	3753
<i>Uncertainty Guardrails for LLM-Based Medical Diagnostic Systems</i>	
Gleb Kuzmin, Lyudmila Rvanova, Ilia Kopanichuk, Vladimir Makharev, Ivan Bakulin, Vladimir Shaposhnikov, Artem Oskin, Petr Anokhin, Aleksandr Nesterov, Artem Vazhentsev, Alexander Panchenko, Elena Tutubalina, Iaroslav Radionovich Besspalov, Dmitry V. Dylov, Artem Shelmanov and Ivan Osledets	3764
<i>Audio Front-End Decoupling: Why Synthetic-Voice Simulators Fail to Predict Real-World Voice Agent Behavior</i>	
Shoutrik Das, Suraj Agrawal and Aashraya Sachdeva	3781
<i>Deploying LLM Agents for Auditable Complaint Investigation in Banking</i>	
John Hartley, Conor Brian Hamill, Azam Din, Fayz Siddiqui, Kingsley Nketia Acheampong, Greig A Cowan and Raad Khraishi	3802
<i>ExhibitionBench: A Multi-Source, Multi-Task Benchmark for Evaluating LLMs as Exhibition Curation Assistants</i>	
Mengze Hong, Sun Yanjie, Mengyao Zhu, Chen Jason Zhang, Di Jiang, Huang Guoqiong Ivanka, Haoyang LI, Zhiqiang Kou, Shuning Zhang, Wangze Ni, Enhao Huang, Zeyang Lei, Xia Zeng, Sheng Wang and Di Liang	3814
<i>CAST: Contrastive Adjudication via Selective Triggering for Clinical LLMs</i>	
Krishna Senthilkumar, Archana Vaidheeswaran and Shivank Garg	3826
<i>TRACE: Deployable Tree-Relational Structure Enhancement for Oncology LLMs</i>	
Jizheng Lai, Yingyun Li, Ying Qin and Haiyang Qian	3836

From Experts to Open-Weight Models: a Validated LLM Pipeline for Institutional Economic News Annotation

Michele Petrocelli, Andrea Rollin, Matteo Berta, Francesca Zafonte, Simone Monaco, Daniele Apiletti and Tania Cerquitelli 3854

When Summaries Distort Decisions: Information Fidelity in LLM-Compressed Financial Analysis

Hoyoung Lee, Suhwan Park, Seunghan Lee, Jun Seo, Jaehoon Lee, Sungdong Yoo, Minjae Kim, CheolWon Na, Zhangyang Wang, Zach Golkhou, Minkyu Kim, Sotirios Sabanis, Alejandro Lopez-Lira, Dhagash Mehta, Soonyoung Lee, Chanyeol Choi, Wonbin Ahn and Yongjae Lee 3870

CapMem: A Benchmark for Caption-Based Episodic Memory in Egocentric Video

Dingli Liang, Yiqiao Xie, Yukai Huang, Zhaokai Wang, Weitong Cai, Guangwen Feng, Jifei Song, Zhensong Zhang and Hang Zhang 3888

ReCAST: Restoration-aware Cascaded Stage-wise Training for Obfuscated SMS Risk Classification

Jieyun Huang, Yi Shen, Kaikai Zhao, Jiangze Yan, Wenjing Zhang, Ping Chen, Ning Wang, Zhaoxiang Liu, Kai Wang and Shiguo Lian 3915

BPCF: A Token-Efficient Process Model Notation for LLM-Based BPMN Generation

Ankit Agrawal, Nicholas Wolf and Flora Geske 3932

Reproducing Omitted Temporal Expressions in Japanese News for Retrieval-Augmented Applications

Tomoaki Yasuda and Shotaro Ishihara 3945

Multi-Step Tool-Calling over Korean Open Public APIs: A Benchmark and a Data-Synthesis Recipe

Dain Kim, Eungi Cho, Kyumin Kim, Shinyeong Noh and Kyuseong Lim 3957

Adapting Technical-Service LLM Agents with Latent Logic Augmentation, Robust Noise Reduction, and Hybrid Reward Modeling

Junzhuo Ma, Chenghuang Shen, Yi Yu, Xingyan Liu, Jing Gu, Hangyi Sun, Guangquan Hu, Jianfeng Liu, Weiting Liu, Pu Mingyue, Wang Yu, Zhengdong Xiao, Rui Xie, Longjiu Luo, Qianrong Wang, Gurong Cui, Honglin Qiao and Wenlian Lu 3987

MMAI Gym for Science: Training Liquid Foundation Models for Drug Discovery

Maksim Kuznetsov, Zulfat Miftahutdinov, Rim Shayakhmetov, Mikolaj Mizera, Roman Schutski, Bogdan Zagribelnyy, Ivan Ilin, Nikita Bondarev, Thomas MacDougall, Mathieu Reymond, Mihir Bafna, Kaeli Kaymak-Loveless, Eugene Babin, Maxim Malkov, Mathias Lechner, Ramin Hasani, Alexander Amini, Vladimir Aladinskiy, Alex Aliper and Alex Zhavoronkov 4023

TabAutoSyn: A Generator-Agnostic Pipeline for Deployment-Ready Synthetic Tabular Data

Dmitrii Kuznechenko, Kirill Rapatskikh, Anna Kalyuzhnaya and Alexander Boukhanovsky 4048

Quantizing Masked Diffusion Language Models: Storage-Quality Trade-offs and Runtime Diagnostics

Pawan Kumar and Neel Mishra 4067

Training Lightweight PII Text Detectors via Rubric-Guided Contrastive Data Synthesis

Shota Ekuni, Jie Zeng, Yohei Kobashi, Takeshi Kojima, Yusuke Iwasawa and Yutaka Matsuo 4090

Verify, Then Argue: Claim Verification in Investment Agent Debates

Joyce Ann Clarize Galang, Ahmed Mady and Georg Groh 4105

GRACE-DS: a Guarded Reward-guided Agent Correction Environment in Data Science

Aleksandr Tsymbalov, Danis Zaripov, Artem Epifanov and Anastasiya Palienko 4120

Hybrid Generative–Tagging Decoding for Product Attribute–Value Extraction

Keshav Kolluru, Manas Gupta, Maroof Ahmad and Suchitra Sathyanarayana 4149

<i>What Else Needs Fixing? Exploring Cost-Effective Test-Time Compute for Revision Propagation in Artifacts Generated Through Conversation</i>	
Daisuke Kikuta	4161
<i>LatentRiskBench: Measuring State-Dependent Risks in Enterprise Agents</i>	
Jingyue Cong, Estrid He, Zhuang Li, Li Shen, Nan Wu, Yameng Peng, Xinshu Li and Andy Song	4178
<i>Certified Cost-Optimal Triage for Suggestion Detection in Reviews</i>	
Aniket Upadhyay, Aakash Trivedi, Dhruv Kumar and Pratik Narang	4196
<i>Fund2Persona: A Framework for Building and Refining Financial Advisor Personas from Fund Disclosure Data</i>	
Suhwan Park, Hoyoung Lee, Zhangyang Wang, Alejandro Lopez-Lira, Young Cha, Chanyeol Choi, Jaewon Choi and Yongjae Lee	4212
<i>Cite or Decline: A Strict Course-Grounded Chatbot for STEM Lecture Videos</i>	
S M Masrur Ahmed and Jaspal Subhlok	4230
<i>When Do Prompt-Side Agent Playbooks Transfer? Accuracy, Cost, and Runtime Shift in Agent Deployment</i>	
Weihong Lin, Xiangzheng Zhang and Lin Sun	4243
<i>Calibrated Deferral Routing for Cost-Efficient Guardrails</i>	
Giandomenico Cornacchia, Inkit Padhi, Manish Nagireddy, Subhajit Chaudhury, Pierre Dognin, Tejaswini Pedapati, Ambrish Rawat, Mark Purcell, Kush R. Varshney and Prasanna Sattigeri . . .	4264
<i>Not All Fallbacks Are Failures: Understanding and Recovering from Fallbacks in Mobile Voice Assistants</i>	
Phillip Schneider, Alexandre Mercier, Joshua Oehms, Kristiina Jokinen and Florian Matthes	4283
<i>AdCompliance: Towards Automated Compliance Verification in Advertising Operations</i>	
Saki Sugiyama, Peinan Zhang, Yusuke Sakai, Sho Takase, Hiroki Ouchi and Taro Watanabe	4295
<i>Reading the News: Adapting Large Language Models to Swedish Journalism Through Continued Pre-Training</i>	
Lukas Borggren, Jenny Kunz and Marco Kuhlmann	4305
<i>MotiGAN: Game Theory Meets the Motivated Intruder Test for Re-identification Risk Generation and Annotation in Pseudonymisation</i>	
Xuan-Son Vu, Tien Dat Dang and Elena Volodina	4328
<i>Cost-Efficient Intent Classification via Semantic-Candidate Adaptation and Selective Capability Routing</i>	
Jiawei Sun and Yongsheng Sang	4338
<i>What Should Be Fixed? Rule-Grounded Corrective Actions for Reworkable Logical Anomalies in Industrial Inspection</i>	
Haitao Zhang, Bosheng Chen, Haote Xu, Xiaotong Tu, Yue Huang and Xinghao Ding	4349
<i>Catching AI-Generated Code at Scale: A Lightweight and Interpretable Detector for the LLM Era</i>	
Arnesh Batra, Krish Thukral, Dev Sharma, Aniket Khandelwal and Arush Gumber	4357
<i>Phoenix-VL 1.5 Medium: Continued Pre-Training of Sovereign Multimodal AI for On-Premises Public-Sector Deployment</i>	
Ze Yu Zhang, Jiale Wang, Jia Hui Bong, Umar Jamil, Wen Jett Lam, Biondi Lee, Eva Lim, Vincent Maladiere, Hong Jin Jonathan Ng, Arka Ray, Elijah Seah, Justin Tan, Fiona Teo, Guang Xiang Teo, Toh Xin Yi Grace, Yeo Shun Yi and Gee Wah Ng	4374

<i>Graph-Guided Selective Unlearning for Language Models: Controlling Support Routes Beyond Forget Seeds</i>	
Waqas Khan, Tabinda Sarwar, Jingyue Cong, Xun Yi and Estrid He	4398
<i>DiffResearch: Accelerating Deep Research with Diffusion Language Models</i>	
Alan-Barsag Gazzaev, Alexey Gavrilov, Mikhail Mozikov and Sergey Muravyov	4415
<i>CARE: Bounded Selective Evidence-to-Judgment Routing for Cost-Aware Document Security Classification</i>	
Yeseul Chang, Woosang Lim, Rahul Kailasa, Simon Shim and Jaewoo Lee	4423
<i>SKILL.state: Scalable Long-Horizon Agent Skills</i>	
Sanket Badhe, Priyanka Tiwari and Jonghyun Chung	4434
<i>Trace-GRAF: Trace-Guided Granularity Routing for Sequential Agentic Workflows</i>	
Anastasiia Linich, Artyom Sosedka and Semen Budenny	4450
<i>Cloud and On-Premises Deployment of Uzbek Legal RAG via Targeted Retriever Fine-Tuning</i>	
Tatul Danielyan, Mariam Avetisyan and Hrant Davtyan	4461