IJCNLP-AACL 2026

The 15th International Joint Conference on Natural Language Processing and the 5th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics

Proceedings of the Conference - Volume 1: Long Paper

November 6-10, 2026

The IJCNLP-AACL organizers gratefully acknowledge the support from the following sponsors.

Co-organizer



Platinum



Gold





Venue Supporter



©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 979-8-89176-439-2

Message from the General Chair

Welcome to AACL-IJCNLP 2026 in Hengqin, China!

This joint conference brings together the Asia-Pacific and international natural language processing communities to share research, exchange ideas, and build connections. We look forward to conversations that connect languages, disciplines, institutions, and different perspectives on the future of our field.

We hope that the conference will provide a welcoming setting for presenting your work, learning from colleagues, and meeting new collaborators. Whether you are joining us in Hengqin or participating online, thank you for contributing to this community.

We thank the organizing committee, reviewers, tutorial instructors, workshop organizers, invited speakers, volunteers, sponsors, and all participants for the work that makes this meeting possible. We wish you a rewarding conference and an enjoyable stay in Hengqin.

Yue Zhang, General Chair
Ruifeng Xu, Ayu Purwarianti, Tong Rong, and Jun Suzuki, Program Chairs

Message from the Local Organizing Chairs

We are delighted to welcome you to the 5th Asia-Pacific Chapter of the Association for Computational Linguistics and the 15th International Joint Conference on Natural Language Processing. It is our great pleasure to host the AACL-IJCNLP community in Hengqin, a dynamic area of the Greater Bay Area just across from Macau.

We sincerely thank the General Chair, Program Chairs, organizing committee members, reviewers, volunteers, sponsors, and all contributors for their dedication and support in making this conference possible. We are also grateful to all authors, speakers, and participants for bringing their work, ideas, and energy to AACL-IJCNLP 2026.

We hope your time at AACL-IJCNLP 2026 will be both productive and enjoyable. Beyond the technical program, we encourage you to experience the local culture and hospitality of Hengqin and the surrounding region, including nearby Macau. We wish you inspiring discussions, fruitful collaborations, and a memorable conference experience.

AACL-IJCNLP 2026 Local Organization Chairs

Derek F. Wong (University of Macau)
Haofen Wang (Tongji University)

Organizing Committee

General Chair

Yue Zhang, Westlake University, China

Program Chairs

Ruifeng Xu, Harbin Institute of Technology, Shenzhen, China
Ayu Purwarianti, Institut Teknologi Bandung, Indonesia
Tong Rong, Singapore Institute of Technology, Singapore
Jun Suzuki, Tohoku University, Japan

Local Organizing Chair

Derek F. Wong, University of Macau
Haofen Wang, Tongji University, China

Local Organizing Committees

Runzhe Zhang, University of Macau
Zhiwen Xie, Central China Normal University, China

Workshop Chair

Xuefeng Bai, Harbin Institute of Technology, Shenzhen, China
Ratish Puduppully, IT University of Copenhagen, Denmark

Tutorial Chair

Zhongqing Wang, Soochow University, China
Naoaki Okazaki, Institute of Science Tokyo, Japan

Demonstration Chair

Ningyu Zhang, Zhejiang University, China
Derry Wijaya, Monash University, Indonesia

Student Research Workshop Chair

Yuxin Cao, Fudan University, China
Shio Hong, Fudan University, China
Yuhang Zhou, Shanghai Innovation Institute, China
Wenhe SUn, Fudan University, China

Best Paper Award Chair

Rui Yan, Wuhan University, China

Publication Chair

Junxian He, Hong Kong University of Science and Technology
Thuy-Trang Vu, Monash University, Australia

Handbook Chair

Leiyang Cui, Tencent AI Lab, China

Publicity Chair

Aditya Joshi, The University of New South Wales, Australia
KyungTae Lim, Korea Advanced Institute of Science Technology, South Korea

Web and App Chair

Xinchi Chen, Fudan University, China
Piji Li, Nanjing University of Aeronautics and Astronautics, China
Shudong Liu, University of Macau

Finance Chair

Hou Pong Chan, Alibaba DAMO Academy

Registration Chair

Wenxuan Zhang, Singapore University of Technology and Design, Singapore

Ethics Chair

Yangqiu Song, Hong Kong University of Science and Technology
Ekaterina Vylomova, The University of Melbourne, Australia

Diversity Inclusion Chair

Mamoru Komachi, Hitotsubashi University, Japan
Emily Ohman, Waseda University, Japan

Student Volunteer Chair

Juntao Li, Soochow University, Japan

Sponsorship Chair

Yan Song, University of Science and Technology of China, China
Yuanhe Tian, Zhongguancun Institute of Artificial Intelligence, China

Keynote Talk

NLP for Mental Health: Risk, Promise, Responsibility

Iryna Gurevych
Technical University of Darmstadt
2026-11-06 09:30:00 – Room: **Muhe B+C**

Abstract: More than a billion people live with a mental health condition, yet care remains scarce and often a poor fit for the person receiving it — so people increasingly turn to unsupervised chatbots to fill the gap. Because language is the richest signal we have into someone's psychological state, natural language processing is the key to doing this better. In this talk, I argue that NLP can genuinely help — but solving it takes clinical grounding, real data, and responsible evaluation, not just larger models. Drawing on work from my group, I show how large language models can turn therapy transcripts into treatment-relevant structure, how clinically grounded synthetic data eases the field's privacy bottleneck, and why evaluation must reflect what clinicians actually care about — not just what's easy to benchmark. I close with an invitation to build mental health AI that is personalized, privacy-aware, and responsibly evaluated: tools that support clinicians, not replace them.

Bio: Iryna Gurevych is a full professor of computer science at the Technical University of Darmstadt, Germany, where she founded and heads the Ubiquitous Knowledge Processing (UKP) Lab. She is a pioneer of computational argumentation and cross-document text modeling, with landmark applications to fact verification and misinformation detection, and her research extends to large language models and interdisciplinary NLP for the social sciences and humanities. Prof. Gurevych served as President of the Association for Computational Linguistics (ACL) in 2023 and is an ACL Fellow, a member of the German National Academy of Sciences Leopoldina and Academia Europaea, and a recipient of the 2025 Royal Society Milner Award.

Keynote Talk

Beyond Self-Play: Towards Effective Self-Evolution in LLMs

Yulan He
King's College London
2026-11-06 14:00:00 – Room: **Muhe B+C**

Abstract: Can large language models keep improving through experience they generate themselves? Self-play and self-generated experience offer promising routes. But producing more synthetic data does not necessarily lead to sustained model capability growth. This talk argues that effective self-evolution depends on whether the learning pipeline continues to provide learnable information. Drawing on our recent work, I will examine several ways such information can arise, including generating informative self-play experience, learning directly from environment-generated experience, and actively seeking novel evidence rather than waiting for tasks to be provided. These studies suggest that effective self-improvement requires more than generating its own experience; it needs experience that contains useful information that the model can actually learn from.

Bio: Yulan He is a Professor in Natural Language Processing at King's College London and a Turing AI Fellow. Her research focuses on making Large Language Models (LLMs) more reliable, interpretable, and safe, particularly in reasoning and agentic AI, with applications spanning across education, health, and science. She has received several awards for her research. Her five-year Turing AI Fellowship project was awarded the Best Research Project (Research Excellence) at the RAi UK AI Robotics Awards 2026. She also received a Best Research Paper Award at the same awards ceremony. In addition, she is the recipient of a SWSA Ten-Year Award and a CIKM Test-of-Time Award.

Keynote Talk

Uncertainty, Belief, and Trust in Language Models

Tim Baldwin
MBZUAI
2026-11-07 09:00:00 – Room: **Muhe B+C**

Abstract: Language models and other AI technologies are being used in an ever-increasing range of applications, including being trialled for decision support or automated decision-making in sensitive applications. A critical component of any such usage is the capability for the model to output a confidence/uncertainty assessment associated with any output, or (statistical) belief underlying a particular action. In this talk, I will present recent work on uncertainty estimation and belief updates, with a particular focus on practical utility in improving model performance and efficiency, and enabling humans to perform tasks more effectively.

Bio: Tim Baldwin is Professor of Natural Language Processing at Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), in addition to being a Melbourne Laureate Professor in the School of Computing and Information Systems at The University of Melbourne, and Co-founder and Chief Scientist of startups including LibrAI (focused on AI safety) and Examn AI (focused on dialogue-based candidate profiling). At MBZUAI he has also held positions including Provost, Associate Provost, and Foundation Chair of the Department of Natural Language Processing.
Tim completed a BSc(CS/Maths) and BA(Linguistics/Japanese) at The University of Melbourne in 1995, and an MEng(CS) and PhD(CS) at the Tokyo Institute of Technology in 1998 and 2001, respectively. He joined MBZUAI at the start of 2022, prior to which he was based at The University of Melbourne for 17 years. His research has been funded by organisations including the Australian Research Council, Google, Microsoft, Amazon, Xerox, ByteDance, SEEK, NTT, and Fujitsu. He is the author of over 500 peer-reviewed publications across diverse topics in natural language processing and machine learning, and the recipient of a number of awards at top conferences.

Keynote Talk

The Genie of the LLM: What Shall We Wish for the Future of NLP?

Nizar Habash
New York University Abu Dhabi
2026-11-08 09:00:00 – Room: **Muhe B+C**

Abstract: Large language models have given NLP extraordinary new capabilities. The genie is out of the lamp. The question is no longer only what it can do, but what we should ask it to do, how we know we got what we really wanted, and what it will do to us, our research, and our education.
I reflect on these questions through a multilingual journey from early work on English, Spanish, and Chinese to decades of Arabic NLP. Arabic provides a revealing case study: its morphology, dialects, scripts, and linguistic variation expose assumptions that broad multilingual experiments can overlook. Recent work illustrates both the remarkable capabilities of LLMs and the continuing value of other models, linguistic knowledge, targeted data, careful evaluation, and human expertise.
The future will not simply be a more powerful version of the present. As answers become easier to generate, good questions, deep understanding, and judgment become more valuable. Our challenge is not merely to build more powerful genies, but to choose our lamps wisely, and become better wish-makers.

Bio: Nizar Habash is a Professor of Computer Science at New York University Abu Dhabi (NYUAD). He is also the director of the Computational Approaches to Modeling Language (CAMeL) Lab. Professor Habash specializes in natural language processing and computational linguistics. He received his PhD in Computer Science from the University of Maryland College Park in 2003. His research includes extensive work on machine translation, morphological analysis, and computational modeling of Arabic and its dialects. Professor Habash has been a principal investigator or co-investigator on over 30 research grants. And he has over 300 publications including a book entitled "Introduction to Arabic Natural Language Processing". Professor Habash is one of the inaugural recipients of the King Salman Academy for Arabic Language Award (2022); he is the recipient of the Antonio Zampolli Prize (2024); and he was selected as a Fellow of the Association for Computational Linguistics (2025). His website is www.nizarhabash.com.

Table of Contents

Towards Probabilistic Question Answering Over Tabular Data
Chen Shen and Estevam Hruschka 1

Generating Linguistically Diverse and Semantically Faithful Patient Reports
Alejandra Lorenzo, Adrien Coulet and Claire Gardent 24

Relational Priors as Convergence Pressure in LLM-Based Multi-Agent Systems
Ming Shen, Chao Shang, Sadat Shahriar, Devang Kulshreshtha, Yi Zhang, Sandesh Swamy and Yanjun Qi 49

Dynamics of Meaning: Towards the Evaluation of Diachronic Semantic Change in Sinhala
Nevidu Jayatilleke and Nisansa de Silva 67

ChartTableViVQA: A Vietnamese Benchmark Dataset for Visual Question Answering with Reasoning Instruction over Charts and Tables Images
Quang Nhat Truong, Pham Quoc Thang and Son T. Luu 99

Navigating the Reality Gap: On-Device Continual Adaptation of ASR for Clinical Telephony
Darshil Chauhan, Adityasinh Solanki, Vansh Patel, Kanav Kapoor, Ritvik Jain, Aditya Bansal, Pratik Narang and Dhruv Kumar 124

SPECI: Fine-grained Evaluation of Specificity for Image Captioning
Son T. Luu, Hiep K. V. Nguyen, Trung Bui, Viet Dac Lai and Le-Minh Nguyen 140

Are Linguistic and Arithmetic Circuits Dissociated in LLMs?
Riku Kisako, Tatsuki Kuribayashi and Ryohei Sasano 157

Steering with Multilingual Lexicons for Sentiment Control in Low-Resource Languages
Lucinda Laurent, Muhammad Dehan Al Kautsar, Alfan Farizki Wicaksono and Fajri Koto . . 174

PRC: An Efficient and Unified Framework for Evaluating Robustness and Controllability of Large Language Models on Few-shot Prompt Variations
Chunpeng Ma 204

JUMPSTARTER: Human-AI Planning with Task-Structured Context Curation
Xuanming Zhang, Sitong Wang, Jenny Ma, Alyssa Hwang, Zhou Yu and Lydia Chilton 223

Decomposing LLM Advice Sycophancy: Content, Style, and Framing Drive Distinct Response Dimensions
Jaya Singh and Verena Kaynig-Fittkau 255

HybridScholar: Personalized Paper Recommendation with Cross Attention-Gated Fusion and Verified Explanation Generation
Joytun Nessa Tonny, Sajidur Rahman, Saddam Mukta, Swakkhar Shatabda and Salekul Islam 275

How Robust Is Multimodal Claim Verification to LLM Rewriting?
Yun-Ang Wu, Xanh Ho, Andre Greiner-Petter, Sunisth Kumar, Tian Cheng Xia, Florian Boudin and Akiko Aizawa 289

Argument Summarization: A Unified Task Definition, Taxonomy, and Survey
Zhiqiang Shi and Oana Cocarascu 307

Escaping the Tokenizer Trap: Probabilistic Semantic Routing for Dense Retrieval in Specialized Domains
Siu Hin Ng 320

ARR-Dv2: New Data from ACL Rolling Review and its Potential to Improve Reviewing in the ACL Community
Atnafu Lambebo Tonja, Sheng Lu, Nils Dycke, Dirk Hovy, Koen Dercksen, Lizhen Qu, Xiaodan Zhu, Thamar Solorio, Iryna Gurevych and Margot Mieskes 333

Attention-Guided Layer Selection for Contrastive Decoding in Large Language Models
Yusuke Sakai, Natthawut Kertkeidkachorn and Kiyoaki Shirai 351

Episodic Memory from Compression Boundaries in Latent Representation Space
David Oneil Campos Ferreira, Priscila Rocha Maia Freitas Ribeiro, Emanuel Borges Passinato, Diogo Fernandes Costa Silva and Arlindo Rodrigues Galvão Filho 364

Protocode: Prototype-Driven Interpretability for Code Generation in LLMs
Bodla Krishna Vamshi and Haizhao Yang 386

North or South? Diagnosing LLMs' Understanding of North Korean Language, Society, and Knowledge
Kiwoong Park, Seyoung Song, Donghyun Woo and Alice Oh 408

HALF: Harm-Aware LLM Fairness Evaluation Aligned with Deployment
Omar El Herraoui, Ali Mekky, Preslav Nakov and Yuxia Wang 427

Sparse Autoencoders Can Capture Language-Specific Concepts Across Diverse Languages
Lyzander Marciano Andrylie, Inaya Rahmanisa, Mahardika Krisna Ihsani, Alfan Farizki Wicaksono, Haryo Akbarianto Wibowo and Alham Fikri Aji 453

Ask Twice: Bidirectional Cross-Encoder Scoring for Faithfulness Evaluation
Arnav Mantro and Leslie Barrett 528

Multi-Perspective LLM Annotations for Valid Analyses in Subjective Tasks
Navya Mehrotra, Adam Visokay and Kristina Gligorić 537

Agent+P: Guiding UI Agents via Symbolic Planning
Shang Ma, Xusheng Xiao and Yanfang Ye 561

Twenty-Five Years of Inter-Annotator Agreement in NLP: A Systematic Survey of Trends, Metrics, and Reporting Practices
Hend Al-Khalifa 579

Auto-AdvER: Disentangling Condition, History, and Commercial Offers in Automotive Advertisements
Filippos Ventirozos, Ioanna Nteka, Tania Nandy, Jozef Bača, Charlie Roadhouse, Peter A. Appleby and Matthew Shardlow 603

Paying for Too Many Tokens? Valid and Cost-Efficient Multimodal LLM Annotation with Simple Heuristics
Zhixi Zhu and Kristina Gligorić 620

DBRAG: Multi-Table Retrieval-Augmented Generation for Complex Database Queries
Prince Larbi Ampofo, Ryoji Kubo and Djellel Difallah 635

BusMA: A Bus Communication Substrate for Multi-Agent Systems
Yanwen Peng, Delvin Ce Zhang, Xi Wang and Nikolaos Aletras 647

DOSA: A Tree-Guided, Self-Regressive Framework for Long Document Structure Analysis
Bohou Li, Benjamin Sowell, Mehul A Shah, Mark Lindblad and Henry Lindeman 670

Supportive Token Revealing for Fast Diffusion Language Model Decoding
Giries Abu Ayoub, Mario Barbara, Lluís Pastor-Pérez, Tanja Bien, Aneesh Barthakur, Alaa Maalouf and Loay Mualem 683

Token-Level On-Policy Contrastive Learning for Black-box Distillation Without Full Rollouts
Jian Yang, Chern Hong Lim and Thuy-Trang Vu 705

A Survey of the Training Process of LLM-Based Agents
Zhisong Zhang, Lemao Liu, Tianqing Fang, Deng Cai and Jiahao Xu 719

CHASE: Adversarial Red-Blue Teaming for Improving LLM Safety using Reinforcement Learning
Rahul Markasserithodi, Aditya Joshi, Yuekang Li, Ishmanbir Singh, Chris Yoo and Alan Niu 742

Iterative Critique-Refine Framework for Enhancing LLM Personalization
Durga Prasad Maram, Dhruvin Gandhi, Zonghai Yao, Gayathri Akkinapalli, Franck Dernoncourt, Yu Wang, Ryan A. Rossi and Nesreen K. Ahmed 759

The Curious Case of Curiosity across Cultures: Evaluating Cross-Cultural Information-Seeking Questions in Humans and LLMs
Angana Borah, Zhijing Jin and Rada Mihalcea 786

EnComp: Lightweight Encoder-Only Context Compression for Retrieval-Augmented Question Answering
Thao Do, Dinh Phu Tran, An Vo, Seon Kwon Kim and Daeyoung Kim 820

Learn To Say NO! Incorporating Model Abstention In Medical Vision-Language Models for Pharmacovigilance
Sofia Jamil, Chinmay Purushottam Bhat, Sooraj Veer R and Sriparna Saha 834

CLFEC: A New Task for Unified Linguistic and Factual Error Correction in paragraph-level Chinese Professional Writing
Jian Kai, Zidong Zhang, Jiwen Chen, Zhengxiang Wu, Songtao Sun, Fuyang Li, Yang Cao and Qiang Liu 849

When the Correct Law Still Fails: A Diagnostic Benchmark for Law-in-Force Retrieval in Versioned Legal RAG
Yuchen Yao, Yiran HU, Kaixiang Zhang, Yixin Su, Jiani Ge and Yun Liu 879

EvidenceRL: Reinforcing Evidence Consistency for Trustworthy Language Models
J. Ben Tamo, Yuxing Lu, Benoit Louis Marteau, Micky C. Nnamdi and May Dongmei Wang 894

AgentArch: A Benchmark for Evaluating Agent Architectures in Enterprise Workflows
Tara Bogavelli, Hari Subramani and Roshnee Sharma 915

CROP: Token-Efficient Reasoning in Large Language Models via Regularized Prompt Optimization
Deep Shah, Sanket Badhe, Nehal Kathrotia and Priyanka Tiwari 934

J-ToRL: Tool-Integrated LLM-as-a-Judge via Reinforcement Learning
Bryan Wilie, Tiezheng YU, Gerry Windiarto Mohamad Dunda, Jierun Chen, Chaofan Tao, Yan Xu, Junxian He and Pascale Fung 952

Comparative Personalization for Multi-document Summarization
Haoyuan Li and Snigdha Chaturvedi 966

Thinking with Deltas: Incentivizing Reinforcement Learning via Differential Visual Reasoning Policy
Shujian Gao, Yuan Wang, Jiangtao Yan, Zuxuan Wu and Yu-Gang Jiang 986

JudgeMoE: Distributional Aggregation for LLM-as-a-Judge
Yiqi Liu, Joseph James, Yang Wang, Kun Zhao, Chenghao Xiao and Chenghua Lin 1016

Method Ranking Inverts Across Targets: A Matched-Grid Evaluation of Jailbreak Attacks on Open-Weight Vision-Language Models
Venkat Aniruddh Thamatam, Vasudev Majhi, Aaryan Nagda, Jagat Sesh Challa, Murari Mandal, Yash Sinha and Dhruv Kumar 1043

The Language of Bargaining: Linguistic Effects in LLM Negotiations
Stuti Sinha, Himanshu Kumar, Aryan Raju Mandapati, Rakshit Sakhuja and Dhruv Kumar . 1064

Exploring Agentic Multimodal Large Language Models: A Survey for AIScientists
Jinglin Jian, Yi R. Fung, Denghui Zhang, Yiqian Liang, Qingyu Chen, Zhiyong Lu and Qingyun Wang 1089

Conversational Intelligent Tutoring Systems: Evolution, Taxonomy, and Pedagogical Frontiers
Zhengyuan Liu, Stella Xin Yin, Li Zhu, Xiaoman Wang, Yunshi Lan and Nancy F. Chen . . . 1134

Reference-Grounded Data Curation for Instruction-Following Thai-English Machine Translation
Thodsaporn Chay-intr, Krittapad Harnchang, Mahannop Thabua, Kobkrit Viriyayudhakorn and Thanaruk Theeramunkong 1152

Are Many Tasks Necessary for General LLM Evaluation? A Matrix Completion View of Benchmark Redundancy
Kohei Tsuji, Yusuke Sakai, Hidetaka Kamigaito and Taro Watanabe 1172

Investigating LLM Capabilities on Long Context Comprehension for Clinical Question Answering
Feras AlMannaa, Talia Tseriotou, Jenny Chim and Maria Liakata 1188

Video-guided Machine Translation for Indic Languages
Pinaki Das, Srikar Ayyagari, Raj Dabre, Thuy-Trang Vu, Virendra Singh, Pushpak Bhattacharyya and Gholamreza Haffari 1214

Lexically Stable but Semantically Improvable: On the Editability of Machine Translation Drafts for LLM Post-Editing
Tianlin Xiao and WenHan Chao 1236

Bahnar–Vietnamese Parallel-Sentence Retrieval: A Low-Resource Case Study with Lexical, Neural, and Hybrid Methods
Thi Ty Nguyen, Long S. T. Nguyen, Phat T. Tran-Truong, Tan Sang Nguyen and Tho Quan . 1260

Atomic and Holistic LLM Judges for Reference-Grounded Support Labels: A Prompt-Controlled Comparison
Xinran Zhang 1281

Learning What Is Said, Not How: Stance-Contrastive Learning for Generalizable Hate Speech Detection
Yucheng Lin, Dan Kienast and Jiaojiao Jiang 1301

Cleaner Speech, Weaker Generalization: Revisiting Pitt-Derived Benchmarks for Alzheimer's Disease Detection
Luqi Sun, Shreeram Suresh Chandra, Lin Zhang, You-Jin Li, Brian MacWhinney, Yu Tsao, Emily Mower Provost and Berrak Sisman 1319

CARE-Bench: Benchmarking Patient-Facing LLM Triage
Yining Hua, Hongbin Na and Cyrus Ayubcha 1340

OpenQTS: A Metadata-Rich Open Dataset for Tang Dynasty Poetry
Robin Young 1364

Geo-spatial and Geo-temporal Reasoning in Vision-Language and Large Language Models : A Review
Shahriyar Zaman Ridoy, Azmine Toushik Wasi, Koushik Ahamed Tonmoy and Md Rizwan Parvez 1377

Knowing When to Critique: Task-Adaptive Metacognitive Regulation for Reliable LLM Reasoning
Xinmeng Hou, Ziting Chang, Zhouquan Lu, Bohao Qu, Liang Wan, Wei Feng, Hai Hu and Qing Guo 1412

Spatial Reasoning in Multimodal Foundation Models: A Survey of Spatial Understanding in LLMs and VLMs
Azmine Toushik Wasi, Md. Masudur Rahman, Abdur Rahman, Mohsin Mahmud Topu, Mohammed Eunus Ali, Riashat Islam and Md Rizwan Parvez 1451

Agent-Environment Alignment via Automated Interface Generation
Kaiming Liu, Xuanyu Lei, Ziyue Wang, Peng Li and Yang Liu 1480

ToneCL:Contrastive Learning for Few-Shot Syllable-Level Tone Classification
Qisheng Liao and Youngah Do 1513

Less Random, More Incoherent: Greedy Hadamard Alignment for LLM Quantization
Yun Ye, Zhaoqi Wu, Xueshu Wang, Cong Zhang, Zhu Yongbo, Zhonghua Sun and Beryl XU1526

DAMG-RAG: Document-Aware Multi-Granularity RAG for Visually Rich Document Understanding
Jiahao Gu, Ningjing Liu, Yuanzhuo Ding and Siheng Chen 1541

AstroAgentBench: Evaluating Agentic Planning on Space Mission Planning Tasks
Weiyi Wang, Xinchi Chen, Jingjing Gong, Xuanjing Huang and Xipeng Qiu 1561

Read the Room Before Your Model Generates
Muyao Kong and Zhihao Lin 1591

Bait-and-Recover: Poisoning Internal Refusal Signals to Defend LLMs against White-Box Editing Jailbreaks
Tian Gao, Zhipeng Xie, Yuhao Wu, Junhua Liu and Xin Fang 1608

T2MBench: Evaluating Text-to-Motion Generation under Diverse Language Instructions
Bin Yang, Rong OU, Weisheng Xu, Jiaqi Xiong, Xintao Li, Taowen Wang, Luyu Zhu, Yuetong Fang, Xu Jiang, Jing Tan and Renjing Xu 1622

State-Anchored Complete-View Distillation for Robust Conversational Multimodal Emotion Recognition
Zhaoyan Pan, Xiangdong Li, Wenke Wu, Mengting Ma, Ye Lou, Ji Zhou, Jiatong Pan and Wei Zhang 1663

Can LLMs Follow the Pulse of a Crisis? Evaluating Crisis Sentiment in Bangladesh's July Uprising
Md. Samiul Alim, Mahir Shahriar Tamim, Tanvir Ahmed Khan, Sharjil Khan, Rafia Ferdous Duti, Shahriyar Zaman Ridoy and Mohammad Ali Moni 1688

Hierarchy Aware Preference Optimization for the Safety of Korean Small Language Models
Carl S. Kim, Junyoung Koh, Kyeonghun Kim and Seunghyeok Hong 1709

CCD-CBT: Multi-Agent Therapeutic Interaction for CBT Guided by Cognitive Conceptualization Diagram
Chang Liu, Changsheng Ma, Yongfeng Tao, Bin Hu and Minqiang Yang 1723

SG-LegalCite: A Principle-Augmented Benchmark for Legal Citation Retrieval in Singapore Law
Lee Yueh Ern Shannon, Kaidong Feng, Yingpeng Du, Lee En Jia Chloe and Zhu Sun 1746

Decoupling Internal Representational Changes and Causal Importance in Fine-Tuned Large Language Models
Lingfang Li, Procheta Sen, Shubham Das and Danushka Bollegala 1767

DeBERTa-ConPara: Attack-Aware and Deployment-Realistic Detection of AI-Generated Text
Mohamed Mady, Yupei Li, Johannes Reschke and Björn Schuller 1792

Cantus: Toward Pedagogical Feedback Generation for Sight-Singing Learning
Jiale Chen, Xuelian Dong, Yubing Chen, Yongxin Hu, Wenxiu Xie and Tianyong Hao 1825

CoSelector: Multi-Agent Collaboration for Multilingual Retrieval-Augmented Generation
Long Yuan, Rui Qi, Xue Zhang, Jinan Xu and Kaiyu Huang 1838

MMA: Multimodal Memory Agent
Yihao Lu, Wanru Cheng, Zeyu Zhang and Hao Tang 1857

Empathy Is Not the Scapegoat for Declines in Factuality and Reliability
Kanwal Mehreen, Letian Shi, Deheng Zhang and Yuxia Wang 1874

RouterKGQA: Specialized–General Model Routing for Constraint-Aware Knowledge Graph Question Answering
Bo Yuan, Hexuan Deng, Long Bang, Xuebo Liu, Zheng Shuai, Zhou Mingzhen and Min Zhang 1903

CoMMA: Conflict-Aware Read-Time Control for Conversational Memory QA
Zixiao Wang, Xingbo Du, Duzhen Zhang and Le Song 1925

LogicLure: Toward Evidence-Grounded Safe Decision-Making under Hidden Risks
Kyungho Kim and JunCheol Shin 1945

Initialisation Determines the Basin: Efficient Codebook Optimisation for Extreme LLM Quantization
Ian W Kennedy and Nafise Sadat Moosavi 1966

FrenchNews-7: Benchmarking Cross-Publisher French News Editorial Desk Classification
Amr Sobhy 1982

SweSpecQA: A Swedish Benchmark for Specialized Document-Native Question Answering
Murathan Kurfali 1998

DuDi: Dual-Signal Distillation with Cross-Lingual Verbalizer
Patomporn Payoungkhamdee, Tinnakit Udsa, Jian Gang Ngui, Sarana Nutanong, Alham Fikri Aji and Peerat Limkonchotiwat 2023

SEA-NLI: Language Inference as a Lens into Southeast Asian Cultural Understanding
Peerawat Chomphooyod, Yosephine Susanto, Jian Gang Ngui, Attapol Rutherford, Alham Fikri Aji, Sarana Nutanong, Can Udomcharoenchaikit and Peerat Limkonchotiwat 2038

Grounding What Users Care About: Multi-Aspect Personalized Modeling from Implicit Preferences
Jiin Park and Misuk Kim 2063

Incorporating Abstract Meaning Representation and Uniform Meaning Representation into In-Context Learning Strategies for Machine Translation
Thu Hoang, Ellie Yang, Weixin Lin and Shira Wein 2090

Beyond Downstream Task Performances: A Set-Theoretic Evaluation of Sentence Embeddings
Yash Mahajan, Naman Bansal, Faria Binte Kader and Santu Karmaker 2125

Are Reasoning Vision-Language Models Robust to Semantic Visual Distractions?
Yizheng Sun, Mochuan Zhan, Yanan Ma, Jia Tong See, Yifan Wang, Ziyi Wang, Hao Li, Yang Cui, Wenhao Cai, Jingyu Sun, Chenghua Lin, Riza Batista-Navarro and Jingyuan Sun 2159

Routing-Based Continual Learning for Multimodal Large Language Models
Jay Mohta, Kenan E. Ak, Gwang Gook Lee, Dimitrios Dimitriadis, Yan Xu and Mingwei Shen 2177

Mechanistic Evidence Routing in Transformers: Margin-Gated Intervention for Retrieval-Conflict Hallucination
Sridurgesh Ravichandran 2193

Prompting Underestimates LLM Capability for Time Series Classification
Dan Schumacher, Erfan Nourbakhsh, Rocky Slavin and Anthony Rios 2212

Universal Boosts, Specific Suppressors: Sparse Autoencoder Steering of Medical Vision-Language Models
Farhad Nooralahzadeh, Benjamin Gundersen, Nicolas Deperrois, Hidetoshi Matsuo, Mizuho Nishio, Thomas Frauenfelder, Ahmed Allam, Christian Bluethgen, Michael Moor and Michael Krauthammer 2241

Language Models Should Be Used to Surface the Unwritten Code of Science
Honglin Bao, Siyang Wu, Jiwoong Choi, Yingrong Mao and James Evans 2264

Absolute Error Is Not Enough: Ranking Model Performance Without Labels to Estimate Disparities at Scale
Veronica Rammouz, Aaron O. Gonzalez, Carlos Cruzportillo, Adrian Tan, Nicole Lang Beebe, Ke Yang and Anthony Rios 2290

Persona Non Grata: LLM Persona-Driven Generations in MCQA are Unstable in Distinct Dimensions
César Guerra-Solano and Xiang Lorraine Li 2308

Large Language Models Approach Expert Pedagogical Quality in Math Tutoring but Differ in Instructional and Linguistic Profiles
Ramatu Oiza Abdulsalam and Segun Aroyehun 2333

Text-Preserving Lossy Text Compression: A Study of Strategic Deletion and LLM Reconstruction
Yuchun Zou, Junhong Tong and Jun Li 2349

INMS: Memory Sharing for Large Language Model based Agents
Hang Gao and Yongfeng Zhang 2383

ATOD: An Evaluation Framework and Benchmark for Agentic Task-Oriented Dialogue Systems
Yifei Zhang, Hooshang Nayyeri, Rinat Khaziev, Emine Yilmaz, Gokhan Tur, Dilek Hakkani-Tür and Hari P Thadakamalla 2399

Embarrassingly Simple Access Control for Large Language Models
Veronica Rammouz, Đorđe Klisura, Ram Krishnan and Anthony Rios 2417

A Unified Ambiguity Signal for Calibrated Multimodal Emotion Recognition
Tahir Ahmad, Raza Mukhtar, Chanjun Park and Yeo Chan Yoon 2432

MisconceptDialogue: Evaluation Criteria Shape Measured Success in LLM Math Tutoring
Woojin Lee, Yoonkyu Woo, Jinah Kim, Hyeok-Min Gwon and Jin-Xia Huang 2454

Tracing the Genesis of Logic: Causal Data Attribution of Mechanistic Features via Counterfactual Poisoning
Xiaokun Wang, Siyu Song, Hongbo Zhao, Wentao Liu, Jiaqi Jiang and Qingqing Mao 2474

Three Numbers and a Precondition: Auditing Multi-Source LLM Benchmarks for Source-Identity Leakage
Tian Xia and Sarp Aykent 2484

PatchHolmes: Agentic Patch Retrieval via Listwise Selection
Guanqun Yang, Yingming Zhou, Jiangrui Zheng, Shudong Hao and Xueqing Liu 2503

Schema-Key Wording as an Instruction Channel in Structured Generation under Constrained Decoding
Yifan Le 2523

Anchored Turns: A Length-Driven Gradient in Chain-of-Thought Causal Necessity Across Multi-Turn Conversations
Dhruv Gupta, Vasudev Majhi, Jagat Sesh Challa, Murari Mandal and Dhruv Kumar 2536

LKValues: Aligning Large Language Models with Sri Lankan Societal Values
Nethmi Muthugala, Supryadi , Surangika Ranathunga, Nisansa de Silva, Ruijie Tao, Ovindu Gunatunga, Pengyun Zhu, Shaowei Zhang, Jingting Zheng and Deyi Xiong 2559

EcoStatKG: A Domain-Specific Statistical Knowledge Graph and Benchmark for Evaluating Numerical Hallucination in LLMs
Ratnesh Kumar Joshi, Sagnik Sengupta and Asif Ekbal 2598

Automatic Evaluation of Mental Health Stigma in Online Communication
Naomi Baes, Jemima Kang, Nick Haslam, Chris Groot, Alsa Wu, Luc Raszewski and Yulia Otmakhova 2622

Domain Knowledge Annealing: A Structured Local-to-Global Paradigm for Constructing Expert LLMs
Yikang Zhang, Zuchao Li, Ping Wang, Lefei Zhang, Hai Zhao and Bo Du 2665

The Geometry of Uncertainty in Language Model Representations
Zy Li 2683

Language Models Detokenize, But Only for Themselves
Qingyun Li and Ruoyun Li 2703

PAIRSAFE: Proxy-Resistant Paired Safety Evaluation for Refusal Boundaries
Wenzhang Du 2720

KaiToM: A Benchmark for Culturally Informed Theory-of-Mind in Thai Context
Kun Kerdthaisong, Pasin Buakhaw, Wayupuk Sommuang and Nutchanon Yongsatianchot . . 2732

Diagnosing Prompt Sensitivity and Interaction Instability in LLM-Based Humor Judgment
Dohyeon Kim and Jihoon Yang 2756

WiC is Not WSD: A Study on LLMs and Lexical Ambiguity Resolution
Yi Zhou, Kiamehr Rezaee, Danushka Bollegala, Mohammad Taher Pilehvar and Jose Camacho-Collados 2799

MM-NeuroOnco: A Multimodal Benchmark and Instruction Dataset for MRI-Based Brain Tumor Diagnosis
Feng Guo, Jiaxiang Liu, Yang Li, Qianqian Shi and Mingkun Xu 2818

MGSM-Pro: A Simple Strategy for Robust Multilingual Mathematical Reasoning Evaluation
Tianyi Xu, Kosei Uemura, Alfred Malengo Kondoro, Catherine Nana Nyaah Essuman, Tadesse Destaw Belay, Ifeoma Okoh, Ganiyat Afolabi, Ayodele Awokoya and David Ifeoluwa Adelani . . . 2846

IndicIFEval: A Benchmark for Verifiable Instruction-Following Evaluation in 14 Indic Languages
Thanmay Jayakumar, Mohammed Safi Ur Rahman Khan, Raj Dabre, Ratish Puduppully and Anoop Kunchukuttan . . . 2864

Rethinking Cross-Modal Attention for Visual Token Pruning in Multimodal LLMs
Diego Weng Qiu and Lin Shang . . . 2886

Corrective Feedback-Driven Humor Generation for Tiny Language Models
Ramakrishna Pinninti, Dibyanayan Bandyopadhyay, Lalit Chandra Routhu, Gaël Dias, Alexis Lechervy, Jeremie Pantin, Mohammed Hasanuzzaman and Asif Ekbal . . . 2898

REMIND: Enhancing CoT Reliability with Self-Supervised Introspection
Souvik Maji, Dibyanayan Bandyopadhyay and Asif Ekbal . . . 2917

DeepVision-103K: A Visually Diverse, Broad-Coverage, and Verifiable Mathematical Dataset for Multimodal Reasoning
Haoxiang Sun, Lizhen Xu, Bing Zhao, Wotao Yin, Wei Wang, Boyu Yang, Rui Wang and HU Wei 2943

SurgAgents: A Multi-agent Framework for Embodied Cataract Surgical Intent Prediction
Shengyuan Wang, Yanzhe Ji, Yue Yao, Zhenyue Qin and Yang Liu . . . 2958

GROWTHEVAL: Evaluating Stage-Conditioned Developmental Realism in Role-Playing LLMs
Danping Yan . . . 2974

S^4R: Selective Sampling, Subspaces, and Sparse Reconstruction for Compressed Long-Context KV Caching
Jialong Han, You Wu and Kewei Tu . . . 2997

QUILT: Query–User Interaction Linguistic Taxonomy
Yuyang Dai . . . 3018

VERGE: Verification-Enhanced Generation of Multi-Hop Evaluation Datasets for Task-Specific RAG
Kyosuke Morita . . . 3036

Encoded but Not Decoded: Layer-Localized Evidence for a Three-Level Gap in LLM Syntax
Zhenyan Lu, He Wang and Xiaohui Huang . . . 3063

Memorization or Extraction? Auditing LLM-Based Recovery of Scientific Assumptions
Suan Lee and JaeSeong Kim . . . 3081

CAFE: Caption-Anchored Feature Enrichment and Evidence Routing for Multimodal Sarcasm Detection
Puchu Zeng, Yun Luo, Qianqian li, Yanzhou Ma, Liu Junyi, Song Fengxuan and Chonghuifang 3103

How Far Do Auto-Interpretation Labels Generalize: A Controlled Study Across Languages, Scripts, and Rewordings
Sripad Karne . . . 3120

The Limits of Simulated Societies: How LLM Alignment Erases Cross-Cultural Variance
Rojin Ziaei . . . 3152

LLM-Ideoplasticity: Measuring Ideological Plasticity in the Political Behavior of LLMs as a Context-Conditioned Distribution
Adib Sakhawat, Syed Rifat Raiyan, Tahsin Islam, Takia Farhin, Hasan Mahmud and Md Kamrul Hasan 3167

ProMoral-Bench: Comparing Prompting Strategies for Moral Judgment and Jailbreak Robustness in LLMs
Rohan Subramanian Thomas, Shikhar Shiromani, Abdullah Chaudhry, Ruizhe Li, Vasu Sharma and Kevin Zhu 3210

Enhancing Safety Alignment by Universal Adversarial Prompts
Yuan Zhou and Xiangyu Zhang 3241

Effects of Theory of Mind and Prosocial Beliefs on Steering Human-Aligned Behaviors of LLMs in Ultimatum Games
Neemesh Yadav, Yihuai Lan, Shan Dong, Mai Hieu Hien, Palakorn Achananuparp, Jing Jiang and Ee-Peng Lim 3252

From Relation-Specific Heads to Shared Interventions: Contextual Entrainment in Language Models
Jinghan Zheng and Jiajun Shi 3278

Edit Policy-Aligned Text Simplification
Xuanxin Wu, Yuki Arase and Masaaki Nagata 3296

Do LLM Embedding Spaces Recover Expert Structure? Evidence from Mental-Health Language
Yixuan Zhu, Zhenke Duan and Fanghen Li 3331

The Model Knows Where to Look: Visual-Grounding-Guided Head Suppression for Mitigating Hallucination in LVLMs
Mong Yuan Sim, Wei Emma Zhang, Xiang Dai and Byron Fang 3347

Prompt-Based Abstention Fails Under Misleading Context: A Controlled Study of Small Frozen RAG Models
Yohanes Andre Setiawan 3368

PersonaPath: Towards Knowledge-Centric Personalized Learning Path Planning
Yu Liu, Zeming Liu, Tianle Zhang, Zihao Cheng, Yuhang Guo, Kehai Chen, Min Zhang, Yunhong Wang and Haifeng Wang 3387

Multicultural Spyfall: Assessing LLMs through a Dynamic Multilingual Social Deduction Game
Haryo Akbarianto Wibowo, Alaa Elsetohy, Qinrong Cui and Alham Fikri Aji 3424

Benchmarking Fine-Grained Readability Explanations for Arabic
Nour Rabih, Khalid N. Elmadani, Hanada Taha and Nizar Habash 3445

Multi-Task Instruction Tuning via Data Scheduling for Low-Resource Arabic SpeechLLMs
Hunzalah Hassan Bhatti, Firoj Alam and Shammur Absar Chowdhury 3473

ROCOB: Dual-Manifold Alignment with Optimal Budgeting for Training-Free PTQ Recovery
Haoqi Yang, Teng Xiao, Zuchao Li, Hao Huang, Lefei Zhang and Hai Zhao 3491

Synergizing Large Language Models and Theory for Human-like Causal Reasoning
Yanxi Zhang, Xin Cong, Zhong Zhang, Xiao Liu, Dongyan Zhao and Yesai Wu 3513

Beyond Isolated Dots: Benchmarking Structured Table Construction as Deep Knowledge Extraction
Tianyun Zhong, Guozhao Mo, Yanjiang Liu, Yihan Chen, Lingdi Kong, Xuanang Chen, Yaojie Lu, Hongyu Lin, Shiwei ye, Shanshan Jiang, Bin Dong, Le Sun, Ben He and Xianpei Han 3545

Explainable End-to-End Argument Mining with Toulmin-Structured Explanations
Zhiqiang Shi and Oana Cocarascu 3575

WebWorld: Web-Simulated World Model
Ting Huang, Zeyu Zhang and Hao Tang 3592

DramaAgent: Agentic Storytelling Video Generation
Ting Huang, Biao Wu, Ronghao Chen, Zeyu Zhang, Tengfei Cheng, Qizhen Lan, Huacan Wang and Hao Tang 3609

You Can Create Words Rather Than Just Detect Them: Towards A New Paradigm of Chinese Neologisms
Qiliang Liang, Hansi Wang, Yue Wang and Yang Liu 3625

From Construction to Injection: Edit-Based Fingerprints for Large Language Models
Yue Li, Xin Yi, Dongsheng Shi, Yongyi Cui, Gerard de Melo and Linlin Wang 3650

Is a Picture Worth a Thousand Words? Adaptive Multimodal Fact-Checking with Visual Evidence Necessity
Jaeyoon Jung, Yejun Yoon and Kunwoo Park 3669

When Token Supervision Fails: Segment-Level On-Policy Distillation for Long-Horizon Agents
Jialong Liu, Zuchao Li, Yuling Shi, Lefei Zhang, Baoyuan Qi and Hai Zhao 3688

Private Evaluation Protocol: Contamination-Resistant Black-Box Evaluation via Public-Slice Surrogates
Md. Mahadi Hassan, Sazzadur Rahaman, Yusuf Sarwar Uddin and Santu Karmaker 3726

Llm agent memory: a survey from a unified representation–management perspective
Zhenheng Tang, Xin He, Tiancheng Zhao, Fanjunduo Wei, Xiang Liu, Peijie Dong, Qian Wang, Qi Li, Huacan Wang, Ronghao Chen, Sen Hu, Weidong Guo, Yu Xu, Haolan Chen, Kunfeng Lai, Kaiyong Zhao, Keyan Ding, Ivor Tsang, Yew-Soon Ong, Bo Li and Xiaowen Chu 3750

Emotion Is Not External Knowledge: Internal Attention Signals for Affective State Reasoning
Wenqiu Tang, Zhen Wan, Yihui Wang, Takahiro Komamizu and Ichiro Ide 3786

IndicRipple: Measuring Cross-Lingual Ripple Effects of Knowledge Editing in Indic Languages
Gunjan Rawat, Sneha Venkatesh, Shashidhar Reddy Javaji and Zining Zhu 3801

A Benchmark for Controllable Speaking-Style Captioning in Audio-Language Models
Shreeram Suresh Chandra, Ashwyn Sharma, Liang Lu, Xiutian Zhao, Leda Sari, Björn Schuller and Berrak Sisman 3820

Hallucinate at the Last in Long Response Generation: A Case Study on Long Document Summarization
Joonho Yang, Seunghyun Yoon, Hwan Chang, Byeongjeong Kim and Hwanhee Lee 3859

What's Worth Asking? Evaluating Preference Elicitation in Transactional Agents
Kandarp Joshi, Tushar Mittal, Chirag Soni, Lillian Zhou and Petar Aleksic 3888

The Same Answer Every Time: Auditing Small Medical LLM Agents
Abhineet Rajendra Kulkarni, Pranav Madhav Kuber and Anurag Bihani 3915

CRAFT: A Unified Counterfactual Reasoning Framework for Tabular Question Answering and Fact Verification
Chenshuo Pan, Yu Zhao, Jie Zhang, Changzai Pan, Jiayi Liang, Yujie Mao, Zhenhe Wu, Shuangyong Song, Yongxiang Li and Zhongjiang He 3934

Multidimensional Consistency Improves Reasoning in Language Models
Huiyuan Lai, Xiao Zhang and Malvina Nissim 3960

CriticGen: Generation-Aware Evaluation as Actionable Feedback
Huifang Du, Zecheng Zuo, Sen Wang, Haofen Wang and Yehui Yang 3975

LLM-Generated Social-Desirability Ratings for Forced-Choice Questionnaire Construction: A Preregistered Downstream Validation
Yui Furukawa, Kyosuke Bunji and Kensuke Okada 3991

What Changes in Hidden States When Model Merging Hurts Task Performance?
Canasai Kruengkrai and Koichiro Yoshino 4021

STRIDE: Structural Reasoning for In-context Demonstration Extraction in Multi-hop Question Answering
Shiwei Chen, Bin Liang, Yue Yu, Kam-Fai Wong, Hui Wang and Ruifeng Xu 4035

ArtHopQA: Diagnosing Hidden-Chain Recovery in Artwork Question Answering
Xuefeng Wei, Xuan Zhou, Yuta Kato, Shintaro Ozaki, Kazuki Hayashi, Yusuke Sakai, Hidetaka Kamigaito, Katsuhiko Hayashi and Taro Watanabe 4057

Role-aware Heuristic Episodic Attention for Conversational LLMs
Hongwanyang, Zhaoning Zhang, Yi Chen, Libo Zhang, Baihui Liu, Linbo Qiao, Zhiliang Tian and Dongsheng Li 4076

AtomicCheck: Learning Structured Reasoning for Fact Verification via Protocol-Guided Reinforcement Learning
Xingquan Zhang, Jianghao Chen, Yang Zhao, Pu Jian, Haoran Gao, Fanyu Meng, Junlan Feng and Chengqing Zong 4092

KoMMERCE: A Korean Multimodal and Multitask Evaluation Benchmark for Reliable Conversational E-commerce Agents
Byunggill Joe, San Kim, Minyoung Jung, Dong Hyok Shin, Yu-Jeong Sung and Saim Shin . 4118

KoEVD: A Korean Benchmark for Metadata-Conditioned Multi-Stage LLM Safety Evaluation
San Kim, Wonjung Kim, Heesoo Kwon and Jae Hee Yang 4151

ConceptWeave: Concept-Driven Prompting through Adaptive Semantic Infilling
Atul Kumar Singh and Rahul Mishra 4190

CulturALL: Benchmarking Multilingual and Multicultural Competence of LLMs on Grounded Tasks
Peiqin Lin, Chenyang Lyu, Wenjiang Luo, Haotian Ye, Md Mehrab Hossain, Chunlan Ma, Shaoxiong Ji, Younes Samih, Bo Zeng, Fan Jiang, Yuanbin Cao, Dilda Duisenbek, Adrian Neo Sau Xun, Pozdniakova Daria, Misevich Liubou, Marinković Nevena, Ngoc Gia Linh Nguyen, Thi Khanh Linh DO, Sarakmatak Sophy, Baotian Hu, Guanhua Chen, Gongbo Tang, Alham Fikri Aji, Longyue Wang and Weihua Luo 4208

Self-Distillation for Multi-Token Prediction
Guoliang Zhao, Ruobing Xie, An Wang, Kai Zhang, Shuaipeng Li, Huaibing Xie and Samm Sun 4225

CLEF: Closed-Loop Error Feedback for Multi-Bit Watermarking in Short LLM-Generated Text
Jing Wang, Kejun Qian, Pengyu Xiong, Qiankun Zhang, Yi Yang and Jing Fan 4244

"Do You Truly Love Me?" Benchmarking LLM Capability on Hierarchical Pragmatic Tactic Conversations
Peter Q. Chen, Pangjing Wu, Zhuohang Jiang, Xiaodong Li and Peter H. F. Ng 4261

What Makes User Reactions Useful for Fake News Detection?
Yassine Toughrai, David Langlois and Smaili 4282

Safety in Batches? Understanding and Mitigating Safety Failures in Batch Prompting
Kihyun Kim, Hee-Seon Kim, Wonjun Lee and Changick Kim 4298

Positive-Unlabeled Contrastive Retrieval for Graduate Program Recommendation
Shiyan Liu and Rui Qu 4320

Does Translation Preserve Character Personality? Quantifying Perceived Character-Personality Consistency Across Languages
Xinger Fu, Masaaki Nagata and Chenhui Chu 4338

Cross-Lingual Semantic Fusion: Towards Universal Sentence Representations via LLMs
Zhen Hu, Zhu Liu and Ying Liu 4359

CGVR: Contradiction-Guided Visual Reasoning for Mitigating Hallucinations in Multimodal Large Language Models
Haoran Ding, Wanqi Zhou, Lei Xing, Lujuan Dang, Kaizhong Zheng and Badong Chen . . . 4375

KBS: A Korean Benchmark Suite for Foundation Model Evaluation
Yeonkyoung So, Jongmin Kim, Sungmok Jung, Gyuseong Lee, Sangho Kim, Jongyeon Park, Joonhak Lee, Seho Pyo, Gyeongje Cho, Seorin Kim, Jisoo Kim, Suyoung Park, Hyunji M. Park, Yelim Ahn, Yeongho Seo and Jaejin Lee 4388

Constructing Disambiguated Knowledge Bases from Large Language Models at Scale
Yujia Hu, Tuan-Phong Nguyen and Simon Razniewski 4414

JanusPO: Asymmetric Clip and KL for Entropy-Stable Group Policy Optimization in Tool-Augmented LLM Reasoning
Hao Li, Man Fung Zhuo, Guangxin Wu, Yusen Wu and Jiayu Wu 4433

Explainable Multimodal Aspect-Based Sentiment Analysis with Dependency-guided Large Language Model
Zhongzheng Wang, Yuanhe Tian, Hongzhi Wang and Yan Song 4451

HorizonBench: Evaluating Ideation Agents on Future Scientific Discoveries
Markus Flicke, Haoyu He, Kashyap Chitta, Yong Cao and Andreas Geiger 4466

A Framework for Evaluating AI-Generated Creativity: Human–LLM Agreement in Evaluating AI-Generated Stories
Pharath Sathya, Yin Jou Huang and Fei Cheng 4480

GLM-RAG: Graph Language Models for Graph-Based Retrieval-Augmented Generation
Maya Arseven, Anette Frank, Beni Egressy, Johann Higl and Moritz Plenz 4503

Structuring the Unstructured: A Systematic Review of Text-to-Structure Generation with a Universal Evaluation Framework
Zheye Deng, Chunkit Chan, Tianshi Zheng, Wei Fan, Weiqi Wang and Yangqiu Song 4533

A2M: Trace-Optimized Agent Hijacking in the MCP Ecosystem
Laizhen Li, Xuan Wang, Peicheng Zhao, Juanjuan Zhao, Kejiang Ye, Cheng-zhong Xu and Xitong Gao 4562

All Verdicts are Not Equal: Rethinking LLM Judge Reliability
Vineet Kumar, Darshita Rathore and Anindya Moitra 4581

MedHal: A Synthetic Dataset for Medical Hallucination Detection
Fabrice Lamarche, Gaya Mehenni, Neshat Elhami Fard, Odette Rios-Ibacache, Li Ming Wang, John Kildea and Amal Zouaq 4599

Structure Tax: How Structured Output affects LLMs Performance
Vineet Kumar, Kanishka and Bhuvanesh Mandora 4620

Grounded or Guessing? LVLM Confidence Estimation via Blind-Image Contrastive Ranking
Reza Khanmohammadi, Erfan Miahi, Simerjot Kaur, Charese Smiley, Ivan Brugere, Kundan S Thind and Mohammad M. Ghassemi 4645

Exploring Cross-Scenario Generality of Agentic Memory Systems: Diagnostics and a Strong Baseline
Zhikai Chen, Jialiang Gu, Junyu Yin, Xianxuan Long, Shenglai Zeng, Xiaoze Liu, Kai Guo, Keren Zhou and Jiliang Tang 4722

Forgetting Only What Matters: Layer-Selective Unlearning toward Robust LLMs
Ravi Ranjan, Olivera Kotevska and Agoritsa Polyzou 4738

A Pre-Training Analogue of Grokking in Language Models: Tracing Delayed Grammatical Generalization
Sherin Muckatira, Namrata Shivagunde, Vijeta Deshpande and Anna Rumshisky 4761

GEAR: Training-Free Rule Distillation for Advanced and Efficient Tool-Augmented Reasoning
Bhiman Kumar Baghel, Fanyou Wu, Suraj Maharjan, Peiran Yao and Srinivasan H. Sengamedu 4781

HARPA: A Testability-Driven, Literature-Grounded Framework for Research Ideation
Rosni Vasu, Peter Jansen, Pao Siangliulue, Cristina Sarasua, Abraham Bernstein, Peter Clark and Bhavana Dalvi Mishra 4821

Retrieve the Logic, Not the Topic: Taxonomic Strategy RAG for Cross-Domain Persuasion Agents
Sana Ayromlou, Purvi Sehgal and Pradyumna Narayana 4881

Fast Listwise Passage Reranking via Semantic Tokens
Omkar Prabhune, Xijun Wang, Nan Bi, Haiyan Wang, Shima Alizadeh, Anoop Jain, Yu Lou and Junbang Liang 4897

TeleSWEBench: A Commit-Driven Benchmark for Evaluating LLM-Powered Software Engineering in Telecommunications
Pranshav Gajjar, Ali Mamaghani, Dinesh Bharadia and Vijay K Shah 4911

Reasoning-Based Personalized Generation for Sparse Users
Bo Ni, Branislav Kveton, Samyadeep Basu, Subhojyoti Mukherjee, Leyao Wang, Franck Dernoncourt, Sungchul Kim, Seunghyun Yoon, Zichao Wang, Ruiyi Zhang, Puneet Mathur, Jihyung Kil, Jiuxiang Gu, Nedim Lipka, Yu Wang, Ryan A. Rossi and Tyler Derr 4926

Identifying User-Inference-Dependent Out-of-Character Behavior in LLM-Based Role-Playing Agents
Shinji Iwata, Shiki Sato, Jun Baba, Koya Ihara and Michimasa Inaba 4927

The Error Is the Signal: Misconception Discovery via Step-Error Annotation Clustering
Robinson Vasquez Ferrer, Mrinmaya Sachan and Shashank Sonkar 4944

LLM-as-RNN: Feedback-Driven Memory Rewriting for Online Learning and Sequential Prediction
Yuxing Lu, J. Ben Tamo, Jinrui Fang, Weichen Zhao, Nan Sun, Xukai Zhao, Yishan Zhong, Wenqi Shi, Jinzhuo Wang and May Dongmei Wang 4977

Epistemic Erosion: An Error Class Beyond Hallucination in LLM Text Generation
Micky C. Nnamdi, Shaunna Wang, Chinaza Nnamdi, Jieru Li, Benoit Louis Marteau, J. Ben Tamo, Yining Yuan and May Dongmei Wang 4998

The bias is in the generator: locating medical chain-of-thought subgroup gaps upstream of the verifier
Micky C. Nnamdi, Benoit Louis Marteau, Shaun Qien Yeau Tan, J. Ben Tamo, Yining Yuan and May Dongmei Wang 5027

Dgrammar: Efficient Constrained Decoding for DLMs
Guan-Ming Chiu and Jeng-Yue Liu 5051

CircuitPile: A Data Synthesis Framework for Enhancing Multimodal LLMs' Perception on Circuit Schematics
Hongyu Chang, Xinyue Li, Huanyong Liu, Feng Zhang and Yuhui Yin 5062

Do Correctness Probes Transfer? A Cross-Family Study of LLM Hidden States
Evan Duan 5095

SWE-Adept: An LLM-Based Agentic Framework for Deep Codebase Analysis and Structured Issue Resolution
Kang He and Kaushik Roy 5119

VTBench: Evaluating Visual Tokenizers for Autoregressive Image Generation
Huawei Lin, Tony Geng, Zhaozhuo Xu and Weijie Zhao 5138

User Feedback to Actionable Insights: A Fine-grained Multilabel Approach for Classifying App Reviews
Chirag Jain and Preethu Rose Anish 5164

Beyond Idealized Patients: Evaluating LLMs under Challenging Patient Behaviors in Medical Consultations
Yahan Li, Xinyi Jie, Wanjia Ruan, Xubei Zhang, Huaijie Zhu, Yicheng Gao, Chaohao Du and Ruishan Liu 5181

Latent Prediction Needs Alignment: A Controlled Study of Joint-Embedding Predictive Vision–Language Learning
Mohammad Kohankhaki, Daniel Kusuma, Shirin Salehi, Carsten Kamp, Sigrid Brell-Cokcan and Anke Schmeink 5210

SaFeR-ToolKit: Structured Reasoning via Virtual Tool Calling for Multimodal Safety
Zixuan Xu, Tiancheng He, Huahui Yi, Kun Wang, Xi Chen, Gongli Xi, Qiankun Li, Kang Li, Yang Liu and Zhigang Zeng 5223

Improving LLM-Based Entity Detection with Annotation Guidelines
Marcel Geromel and Philipp Cimiano 5251

Do not be greedy, Think Twice: Sampling and Selection for Document-level Information Extraction
Mikel Zubillaga, Oscar Sainz, Oier Lopez De Lacalle and Eneko Agirre 5269

RL-ARC: Calibrating Large Reasoning Models via Reasoning-guided Uncertainty
Gukhyeon Lee and SangKeun Lee 5284

HeRE: A Novel Benchmark for Hebrew Relation Extraction via LLM-Guided Denoising of Knowledge Graph Alignments
Ron Keinan, Amir David Nissan Cohen and Reut Tsarfaty 5311

Query-Conditioned Graph Retrieval for Contextualized LLM Reasoning in Personalized Wearable Data
Zhenyu Lu, Mahyar Abbasian and Amir M. Rahmani 5330

Policy-Conditioned Constrained Decoding for Column-Level Access Control in Text-to-SQL
Miyamoto Ryoto, Xin Fan and Hayato Yamana 5377

A Systematic Review of Meta-Reasoning for Large Reasoning Models
Haoran Yu, Zilong Wang, Kang Pan, Xiao Zhang, Erli Meng and Shuai Li 5392

Domain-Agnostic Neural Topic Modeling with Contextual Token-Level Semantic Graph Representation
Seung-Won Seo, Won Ik Cho and Yongmin Yoo 5415

Surfacing Minor Nuances in Subjective Expressions: Initial Evidence for Improved Output Quality in Collaborative Caption Revision
Takeru Isaka, Jiro Nagao and Iwaki Toshima 5430

PoliSpace: Political Opinion Prediction using Shared Latent Space
Johanna D'Ciofalo Khodaverdian, Frédéric Berdoz and Roger Wattenhofer 5447

Predicting Mergeability of Language Model Domain Experts
Ervin Dervishaj, Quynh Do, Giannis Karamanolakis, Anna Rumshisky, Salma Sohrabi-Jahromi and Judith Gaspers 5464

RAGCap-Bench: Benchmarking Capabilities of LLMs in Agentic Retrieval Augmented Generation Systems
Jingru Lin, Chen Zhang, Yong Liu and Haizhou Li 5478

Beyond Retrieval: Deterministic Memory Governance with Pairwise Constraints and Certified Repair for Long-Term Language Agents
Panayiotis Christou, Miguel Palma, Zefan Du, Kaixun Hua and Ying Mao 5501

HDSL: A Hierarchical Domain-Specific Language for Structured 3D Indoor Scene Generation and Localized Editing with LLM Agents
Letian Li, Chao Shen, Shuzhao Xie, Chenghao Gu, ZhengXiao He, Yu Meng, Xin Yang, Wenyuan Jiang and Zhi Wang 5519

Training LLMs Beyond Next Token Prediction - Filling the Mutual Information Gap
Chun-Hao Yang, Bo-Han Feng, Tzu-Yuan Lai, Yan-Yu Chen, Yin-Kai Dean Huang and Shou-De Lin 5539

Structural vs Cultural Reasoning: Reliability Limits of LLM-as-a-Judge in Multilingual Evaluation
Shehenaz Hossain, Houda Bouamor and Haithem Afli 5555

Decisions by Discrimination: Differential Evidence for Medical Reasoning
Yunpeng Xiao, Yuzhang Xie, Junhao Huang, Carl Yang, Xiao Hu and Kai Shu 5589

The Expense of Seeing: Attaining Trustworthy Multimodal Reasoning Within the Monolithic Paradigm
Karan Goyal 5604

Ontology-Guided Knowledge Graph Construction from Structured Metadata and Climate Science Publications
Aayush Acharya, Qi Zhang, Eduard Dragut and Longin Jan Latecki 5615

Sampling via Decision-Flow: Training-Free Extraction of Improved Latent Reasoning Paths in Large Language Models
Zhendong Mi and Shaoyi Huang 5643

Does Fairness Survive Model Adaptation in Legal Judgment Prediction?
Sarah Kamoun, Parth Patel and Faezeh Ensan 5653

Activation-Efficient Fine-Tuning via Sequence-Compressed Low-Rank Subspaces
Irene Tenison, Stella Ahn, Miriam Kim, Ebtisam Alshehri and Lalana Kagal 5668

Detect Before You Disambiguate: Perturbation-Based Input Ambiguity Detection in LLMs
Tiejin Chen, Xiaoou Liu, Ahmadreza Moradipari, Michael Clifford, Kyungtae Han, Yitao Chen, Nejib Ammar and Hua Wei 5688

Diverse Reasoning Paths Matter: Symbolic Equivalence Augmentation for Enhancing RL Exploration
Xiaoou Liu, Tiejin Chen, Hanyang Chen, Yuan Ling and Hua Wei 5703

Information-Theoretic Design of Sparse Autoencoder-Based Steering
Hirofumi Morise 5717

An Empirical Study of Position Bias in Visualized Text Retrieval
Cuiqiaoshu Pan, Ziyang Zeng, Dun Zhang and Yuqing Yang 5740

A Geometric Perspective on Stabilizing Value Conflict Resolution
Saket Reddy and Andy Liu 5761

Do LLMs Have Social Sense? Evaluating Alignment of Conversation Understanding and Reasoning
Zhaoqing Wu and Dan Goldwasser 5786

Encoded but Not Express: Probing Subtle Speech-Conditioned Emotion in Large Audio Language Models
Jingxi Liang, Pengze Guo, Zhiwen Xie, Qifeng Wang, Yujia Huo and Derek F. Wong 5809

See Every Block Once: Variance-Reduced Post-Training for Masked Diffusion Language Models
Yunzhen He, Hiroaki Yamagiwa and Hidetoshi Shimodaira 5836

Amortizing Self-Correction for Continual Embodied Adaptation
Woo Kyung Kim, Jooyoung Kim, Minjong Yoo and Honguk Woo 5858

Back to Basics, Powered by Data: Character-Level Transformers for Robust Chinese Machine-Generated Text Detection
Shengjie Xu, Zhenghao Gao, Rujie Dai, Zijing Li, Xiaoyang Xie and Zehua Lyu 5879

PULSE-CARE: A Measurement-Validity Audit of Simulated-Patient Clinical Agent Evaluation
Srimonti Dutta, Dinky Mishra, Aditi Singh and Akshata Kishore Moharir 5904

DensePath REINFORCE: Lightweight Policy Improvement from SFT-Induced Token-Level Reward Signals
Jiangnan Li, Thuy-Trang Vu, Ehsan Abbasnejad and Gholamreza Haffari 5924

Not All Tokens Are Equally Important: Auditing and Compressing Redundancy in LLM Reasoning
Nguyen-Khang Le, Dieu-Hien Nguyen, Truong Dinh Do and Le-Minh Nguyen 5942

ICCU: In-Context Continual Unlearning via Pattern-Induced Refusal Rules
Ruihao Pan and Suhang Wang 5963

PerQ: Efficient Evaluation of Multilingual Text Platform-Adaptation Quality
Dominik Macko and Andrew Pulver 5990

BrainMoE: A Brain-Inspired Modular Mixture-of-Experts Model for Multimodal Reasoning
Siyao Song, Yunhao Zhang, Jingyuan Sun, Yaping Zhang and Chengqing Zong 6004

DALSA: Difficulty-Aware Long-to-Short LLM Reasoning with Model Merging
Dongwook Min, Jeonhwi Kang, Donghyeon Jeon, Inho Kang and Seung-Hoon Na 6023

From Preferences to Scores: Semi-Supervised Multi-Trait Essay Scoring with LLM Pairwise Preferences and Few Human Anchors
Takumi Shibata and Yuichi Miyamura 6041

Orthogonal Style Injection: Mechanistic Persona Steering for Zero-Shot Cross-Lingual Subtitle Translation
Tarun Chintada, Kshetrimayum Boynao Singh and Asif Ekbal 6058