EMNLP 2026

The 2026 Conference on Empirical Methods in Natural Language Processing (EMNLP 2026)

Proceedings of the Conference

October 24-29, 2026

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

Diamond Level





Platinum Level



















Gold Level









Tencent tether.

Silver Level

Adobe

Bronze Level

DATAmundi.ai DataoceanAI translated. Booking.com

ALICE Kenable 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 979-8-89176-459-0

Message from General Chair

Welcome to the 31st edition of the **Conference on Empirical Methods in Natural Language Processing (EMNLP)**!

EMNLP is organized annually by SIGDAT, the Association for Computational Linguistics' special interest group devoted to linguistic data and corpus-based approaches to NLP. This year, we are delighted to bring the conference to **Budapest, Hungary, from October 24–29**. It is a particularly special edition of EMNLP: the largest in the conference's history, with more than **17,000 submissions, over 5,000 accepted papers, and thousands of researchers, students, practitioners, and friends of the community** coming together from around the world.

This scale is a remarkable reflection of the vitality of our field—but it also brings extraordinary organizational challenges. Making EMNLP possible at this size requires an enormous amount of coordination, care, and behind-the-scenes work. The conference would not have been possible without the outstanding support of the ACL professional conference organization team—**Jenn Rachford, Lina Staggs, Hannah Pietsch and Damira Mrsic with Underline**—as well as the many volunteers who worked tirelessly across countless tasks, from coordinating communications and room assignments to managing acceptance decisions, administrative and financial matters, and the inevitable last-minute requests and unexpected challenges that arise when bringing thousands of people together.

A conference of this magnitude can only succeed through collaboration. Organizing EMNLP has been a wonderful opportunity to work with an exceptional group of people, and I have learned a great deal from all of them. I am deeply grateful to every chair, committee member, reviewer, and volunteer whose dedication has turned months of planning into the conference we are now able to share.

Our **Program Chairs—Hung-Yi Lee, Sunipa Dev, Jessy Li, and Desmond Elliott**, together with **Senior Program Chair Christos Christodoulopoulos**—have done an extraordinary job assembling an exciting and diverse scientific program. The number of submissions exceeded even our most ambitious expectations, while the reviewing process brought a number of new challenges, including AI-generated submissions, contribution slicing, and hallucinated references. The program team had to respond to many of these challenges in real time, developing new approaches while coordinating the largest program committee EMNLP has ever had and managing an exceptionally demanding reviewing and decision-making process.

They also took the opportunity to explore new approaches to peer review through the **AI Reviewing Experiment**, opening an important conversation about how emerging technologies may shape the future of scientific publishing and evaluation.

The scientific program is complemented by three outstanding keynote speakers—**Pascale Fung, Graham Neubig, and Anna Korhonen**—who will share exciting perspectives and research with our community. We will also have the opportunity to look beyond the immediate challenges of today through the panel **"New Missions in NLP,"** exploring the future directions of our field and the many scientific and societal opportunities that lie ahead.

And this year, our discussions take place in **Budapest**, a city with a remarkable history at the crossroads of cultures, languages, and ideas. From the banks of the Danube to the historic neighborhoods of Buda and Pest, Budapest offers a fitting setting for a conference devoted to language, communication, and the exchange of knowledge. I hope that, alongside the talks, papers, workshops, and conversations, you will have the chance to experience the city itself—its architecture, cafés, food, music, and vibrant intellectual and cultural life. Perhaps EMNLP 2026 will be remembered not only for the research we shared, but also for the conversations and friendships that began here.

My heartfelt thanks go to the entire organization committee for their extraordinary work and tireless commitment. A conference of this scale simply would not exist without them. If you meet any of the people listed below during the conference, please take a moment to thank them for everything they have done.

- **Program Chairs:** Hung-Yi Lee, Sunipa Dev, Jessy Li, Desmond Elliott

- **Senior Program Chair:** Christos Christodoulopoulos
- **Visa/Local Organizing Chairs:** Noémi Ligeti-Nagy, Veronika Lipp
- **ARR May 2026 Editors-in-Chief:** Giovanni Da San Martino, Jing Jiang, Xiaojun Wan, Michael White, Xiaodan Zhu
- **Award Chairs:** Ivan Titov, Luciana Benotti, Nancy Chen
- **Industry Track Chairs:** Pepa Atanasova, Gholamreza Haffari, Minghan Wang
- **Workshop Chairs:** Chrysoula Zerva, Steffen Eger, George Filandrianos, Jun Suzuki,
- **Tutorial Chairs:** Goran Glavas, Ana Marasovic
- **Virtual Infrastructure Chairs:** Fernando Batista, Ali Hurriyetoglu
- **Sponsorship Chairs:** Nikos Aletras, Matthias Gallé
- **Ethics Chairs:** Anil Ramakrishna, Giuseppe Attanasio
- **Demonstration Chairs:** Sweta Agrawal, Alberto Abad, Lingpeng Kong
- **Publication Chairs:** Shaoxiong Ji, Junjie Hu
- **Handbook Chairs:** Vlad Niculae, Jia Xu
- **Publicity Chairs:** Yuki Arase, Eran Hirsch
- **Student Volunteer Chairs:** Nafise Sadat Moosavi, Favour Aghaebe
- **Diversity and Inclusion Chairs:** Anne Lauscher, David Ifeoluwa Adelani, Spandana Gella
- **Website Chairs:** Raj Dabre, Hour Kaing
- **Workflow Chairs:** Cheng-Han Chiang, Yating Wu

I would also like to extend my sincere thanks to the **Area Chairs, Senior Area Chairs, and the countless reviewers** whose time and expertise made the scientific program possible. My thanks go as well to our local chair **Noémi Ligeti-Nagy** and to the **SIGDAT Board—Kai-Wei Chang, Raquel Fernández, Yue Zhang, and Jose Camacho-Collados**, as well as former officers **Alice Oh and Juan Pino**, for their support, guidance, and trust throughout the organization of the conference.

A special word of thanks goes to **David Yarowsky**, one of the founders of EMNLP, and **Yuji Matsumoto**, whose perspectives on the history of the conference and continued support help connect this 31st edition with the community and ideas from which EMNLP grew.

Finally, I hope that EMNLP 2026 will be more than a successful scientific meeting. I hope it will be a place where new ideas emerge, collaborations begin, old friendships are renewed, and new ones are formed. Most of all, I hope you leave **Budapest inspired by the research, energized by the people you met, and motivated to return to your work with new questions and new ideas**.

Thank you for being part of EMNLP 2026. I hope you enjoy the conference—and your time in Budapest!

André Martins
EMNLP 2026 General Chair

Message from Program Chairs

Welcome to the 2026 Conference on Empirical Methods in Natural Language Processing! We look forward to joining you in Budapest, a city with a rich history stretching back to the Celtic times. We are excited to present a diverse program featuring over 5,000 accepted papers distributed across the Main Conference, Findings, the Industry Track, the Demo Track, and the 28 Workshops. In addition to this exciting scientific program, we hope you will have the opportunity to take a stroll along the Danube river, enjoy some of the local delicacies, and perhaps even visit one of the famous thermal spas. The organization of this conference has been a huge collaborative effort involving hundreds of people. We would like to acknowledge the support of the following people:

- General Chair, André F. T. Martins
- Senior Program Chair, Christos Christodoulopoulos
- The SIGDAT Executive Officers, Kai-Wei Chang, Raquel Fernández, and Yue Zhang, and the former SIGDAT President Alice Oh
- The ARR Editors-in-Chief of the May 2026 cycle: Giovanni Da San Martino, Jing Jiang, Xiaojun Wan, Michael White, and Xiaodan Zhu
- The ARR technical staff: Holy Lovenia and Yuntian Deng
- The Workflow Chairs, Cheng-Han Chiang, Yating Wu, and Mike Zhang
- The OpenReview team
- The 287 Senior Area Chairs
- The 1,445 Area Chairs and the 11,418 reviewers
- Jennifer Rachford and the conference support staff
- Damira Mrsic and Underline team
- The Awards Committee Chairs, Luciana Benotti, Nancy Chen, and Ivan Titov, and the awards committee members
- The Ethics Chairs, Anil Ramakrishna and Giuseppe Attanasio
- The Industry Track Chairs, Pepa Atanasova, Gholamreza Haffari, and Minghan Wang
- The Demonstration Chairs, Lingpeng Kong, Sweta Agrawal, and Alberto Abad
- The Website Chairs, Raj Dabre and Hour Kaing
- The Publication Chairs, Shaoxiong Ji and Junjie Hu
- The Handbook Chairs, Vlad Niculae and Jia Xu
- The Local Organization Chairs, Veronika Lipp, Noémi Ligeti-Nagy, Brigitta Ivanyi-Bitter, and their team
- The Publicity Chairs, Yuki Arase and Eran Hirsch
- The Student Volunteer Chairs, Nafise Sadat Moosavi and Favour Y Aghaebe, and all the volunteers
- The Diversity/Inclusion Chairs, Anne Lauscher, David Ifeoluwa Adelani, and Spandana Gella

- The Virtual Infrastructure Chairs, Fernando Batista and Ali Hurriyetoglu
- The Workshop Chairs, Chrysoula Zerva, Steffen Eger, and Jun Suzuki
- The Tutorial Chairs, Goran Glavas and Ana Marasović
- The Sponsorship Chairs, Nikos Aletras and Matthias Gallé
- The AI Reviewing Experiment Chairs, Joydeep Biswas and Gaurav Sahu
- The ACL Anthology Director Matt Post and the team
- The TACL editors-in-chief Asli Celikyilmaz, Yun-Nung (Vivian) Chen, Ehud Reiter, and Dilek Hakkani Tur, and the Editorial Assistant Yesari Tolga Eden
- The CL Editor-in-Chief, Wei Lu, and the Editorial Assistant Cynthia-Anne Robinson
- All the authors of papers submitted for review

Review Process

All submissions to EMNLP 2026 went through a two-stage review process. First, papers were submitted to the ACL Rolling Review (ARR), where they were reviewed by reviewers and received meta-reviews from area chairs. Authors then had the option to commit their reviewed papers to EMNLP via a separate EMNLP 2026 commitment site. At this stage, senior area chairs provided recommendations, and final acceptance decisions were made by the program chairs. This process is consistent with previous conferences. We worked closely with the ARR team, particularly the Editors-in-Chief for the May 2026 cycle who served as guest Editors-in-Chief for this round, and the ARR technical staff. We recruited additional reviewers and area chairs for ARR, bringing the total to 11,418 reviewers and 1,445 area chairs in the May 2026 ARR cycle, who handled the majority of EMNLP 2026 submissions. We also recruited 287 senior area chairs to oversee the review and meta-review processes, and to provide final recommendations during commitment stage.

Acceptance Rate

The acceptance rate calculation follows precedent set by previous conferences that also go through ACL Rolling Review (ARR). The calculation takes into account the multi-stage process of ARR, where a paper may get revised in ARR and then later committed to the conference. The denominator includes all those committed to EMNLP from ARR February 2025 through ARR May 2026.[1] Among these submissions, 2,033 were withdrawn, either before or after receiving three reviews, and 1,422 were desk-rejected, which are still counted in the denominator. In total, there are 2,704 papers accepted to the Main Conference and 2,478 papers accepted to Findings. The final acceptance rate for Main Conference papers is therefore: 2704/17661 = 15.31%. The Main Conference proceedings will include 2,626 long papers and 78 short papers. Findings papers are those which are not accepted at the Main Conference, but nevertheless have been judged worthy of publication as "solid work with sufficient substance, quality and novelty". We accepted 2478/17661 = 14.03% of papers to the Findings of the ACL: EMNLP 2026. The Findings proceedings will include 2,380 long papers and 98 short papers.

New Initiative: Paper Integrity Policy

We introduced a new Paper Integrity Policy to address the growing concerns in the community regarding irresponsible AI use in paper submissions that overwhelm community resources and make marginal contributions to the field.

[1] For the curious reader, we note that 8,642 papers were committed from ARR May 2026.

	Long	Short
Papers from ARR May 2026	15,581	1,102
Papers from ARR March 2026	706	36
Papers from ARR January 2026	194	14
Papers from ARR October 2025	17	2
Papers from ARR May 2025	3	1
Papers from ARR July 2025	1	2
Papers from ARR February 2025	1	1
Total	16,503	1,158

Thinly Sliced Contributions. This policy enforces the new ARR originality requirement that "any concurrently submitted papers on a related topic with an overlapping set of authors must cite each other and discuss the differences in the related work section". Outside of ARR and EMNLP, we coordinated with conferences with overlapping submission windows (COLM, NeurIPS and AAAI); dual submissions, or thinly sliced submissions that do not follow the originality requirement, were desk rejected. In the end, we desk rejected 8 submissions that overlapped with COLM, 40 submissions that overlapped with AAAI, and 32 submissions that overlapped with NeurIPS.

Unverifiable References EMNLP 2026 partnered with GPTZero to detect references that do not exist in academic databases in all submitted papers.[2] The process for desk rejecting papers followed two routes: (i) reviewers or ARR Area Chairs reported unverifiable references themselves; or (ii) unverifiable references were automatically identified by GPTZero and manually checked by ARR Area Chairs. Papers were desk rejected if references broke either of these criteria: (a) there was at least one reference to a paper that did not exist with the given title; or (b) at least three references had major errors in their author lists, where $\geq 50\%$ of the authors did not match the expected reference. The final decisions were made by either an EMNLP Program Chair or ARR May 2026 Cycle Editor-in-Chief. In total, we rejected 258 submissions written by 1,166 authors. The authors were also informed about the reference(s) that caused their paper to be desk rejected. We sanctioned all 1,166 authors by preventing them from committing their desk rejected papers to EMNLP 2026. We will also recommend an additional sanction to the EMNLP 2027 Program Chairs to prevent the 35 authors with multiple papers desk rejected under this policy from committing any papers to the EMNLP 2027 conference. These sanctions were discussed with and approved by the SIGDAT Executive Leadership. Desk rejected submissions have also been forwarded to the ACL Publications Ethics Committee. This process was repeated for the Camera Ready submissions.

New Initiative: AI Review Experiment

EMNLP 2026 ran an AI Reviewing Experiment to collect feedback from authors about the quality of AI reviews of their submissions. This experiment took place in an opt-in basis, where authors needed to declare whether they agreed to take part in the experiment. We collaborated with two research groups to provide AI-generated reviews for opted-in submissions:

- The ReviewerToo system from researchers at Mila - Quebec AI Institute, ServiceNow Research, HEC Montréal, Polytechnique Montréal, Université de Montréal, and Unified Sciences. This system used two different backbone LLMs: GPT-OSS-120B and GLM 4.7 Flash.
- The AAAI AI Review System, developed by a multi-institutional research team consisting of researchers from the University of Texas at Austin, the University of Alberta, the University of Mi-

[2] More details in the GPTZero technical report: https://gptzero.me/news/how-does-gptzeros-hallucination-check-detector-work-a-technical-report/

chigan, and Oregon State University. This system used the GPT-5.4 under a Zero Data Retention agreement to protect the privacy of the submissions.

In total, 7,921 submissions opted-in to the experiment and received an AI review of their paper. After removing 1,407 papers that were either desk rejected or withdrawn by the authors, we collected quantitative and qualitative feedback from the authors of 1,282 papers. This represents a 1282/6514=19.68% response rate, after removing desk rejected submissions. We also collected feedback from the authors of 921 opted-out submissions, representing a 10.5% response rate. We will present the preliminary findings of this experiment on Sunday, October 25 — Main Conference Day 1 in the Plenary Room from 14:00–14:30.

Special Theme: New Missions for NLP

Large language models have rapidly shifted from research prototypes to widely used infrastructure. This new reality is changing the center of gravity of NLP research, and progress can no longer be measured only by fractional improvements in benchmark scores. When strong general-purpose models are increasingly available through open releases and commercial platforms, the field faces a more foundational question: What are the missions of NLP research now, and what kinds of contributions most advance that mission?

The EMNLP 2026 special theme invited papers that articulate, test, and advance visions for NLP in this era. We welcome work that reflects on the mission(s) of NLP research, reframes our collective research goals, proposes new evaluation and scientific methodologies, and builds bridges across multiple disciplines that can sharpen both our theories and our impact. Submissions may be empirical, theoretical, or position and survey papers, but should be grounded in clear claims, strong evidence, and actionable insights. We received 322 submissions, of which 45 were accepted to the Main Conference and 11 were accepted to Findings.

Best Paper Selection

This year, we received an exceptional collection of papers with novel and notable contributions. To highlight the most outstanding work, we asked our reviewers, ACs, and SACs to all nominate papers they thought were the strongest contributions to the field. In parallel, the Program Chairs identified an additional selection of standout manuscripts. This consolidated set of nominated papers was then forwarded to our Best Paper Award Committee chaired by Ivan Titov, Luciana Benotti, and Nancy Chen. A committee of 49 members conducted a rigorous, comparative review of the anonymized camera-ready papers. We will announce the award winners on Tuesday, October 27 - Main Conference Day 3 in the Plenary Room from 16:00–17:00. It will be our pleasure to honour the Best Papers and Outstanding Papers of EMNLP 2026 in this session.

Program Composition & Presentation Modes

This year, 490 papers were selected for oral presentation by the Program Chairs, with the goal of creating a well-rounded program featuring a diverse set of topics. We placed a high value on representation across tracks and topical coherence in the sessions. In addition to the Main Conference papers, the program also includes 13 papers accepted by Computational Linguistics and 49 papers accepted by Transactions of the ACL (TACL), of which 28 will be presented as in-person oral presentations, thematically distributed across sessions, according to author preferences. Main Conference and journal papers that were not selected for oral presentation will be presented as a poster alongside in-person presenters of Findings papers in designated sessions. Finally, in a change from previous events, in-person presenters are only required to upload materials to Underline that correspond to their presentation type. The Virtual Conference will take place in (mostly) parallel sessions alongside the main conference. This decision is based

on discussions with the General Chair, Jennifer Rachford's team and the Underline team led by Damira Mrsic. There are two virtual conference sessions per day, in an effort to accommodate virtual attendees from across the globe. We have distributed the keynote talks throughout the program in order to support increased attendance by virtual conference attendees. Rounding out the program is a panel discussion, a session reporting the findings of the AI Reviewing experiment, a business meeting, dedicated sessions for the demonstrations track, and Birds-of-a-Feather sessions.

Keynotes and Panel

We are delighted to welcome three keynote speakers to EMNLP 2026, offering complementary perspectives on the future of AI and natural language processing:

- **Pascale Fung (AMI Labs and The Hong Kong University of Science and Technology)** will present "Towards AI that understands the Real World."
- **Graham Neubig (Carnegie Mellon University and OpenHands)** will present "On the Value of Open Research in Language Modeling - An Adversarial Dialog."
- **Anna Korhonen (University of Cambridge)** will present "AI for the Many: Starting from the World, Not the Model."

Our program also features a panel on this year's special theme, "New Missions for NLP Research", moderated by **Raymond J. Mooney** (The University of Texas at Austin), with contributions from **Dirk Hovy** (Bocconi University), **Lillian Lee** (Cornell University), and **Mor Geva** (Tel Aviv University).

We thank our keynote speakers, panel moderator, and panelists for sharing their perspectives with the EMNLP community. We hope that you will enjoy this year's program and hybrid conference!

Sunipa Dev, Google
Desmond Elliott, University of Copenhagen
Hung-yi Lee, National Taiwan University
Junyi Jessy Li, The University of Texas at Austin

Organizing Committee

General Chair

André Martins, Instituto Superior Técnico, University of Lisbon

Senior Program Chair

Christos Christodoulopoulos, Information Commissioner's Office

Program Chairs

Sunipa Dev, Google
Jessy Li, UT Austin
Hung-Yi Lee, National Taiwan University
Desmond Elliott, University of Copenhagen

Local and Visa Chairs

Veronika Lipp, ELTE Research Centre for Linguistics
Noémi Ligeti-Nagy, ELTE Research Centre for Linguistics
Brigitta Ivanyi-Bitter,

Workshop Chairs

Chrysoula Zerva, Instituto Superior Técnico, University of Lisbon
Steffen Eger, University of Technology Nuremberg
Jun Suzuki, Tohoku University

Tutorial Chairs

Goran Glavaš, University of Würzburg
Ana Marasović, University of Utah

Industry Track Chairs

Pepa Atanasova, University of Copenhagen
Gholamreza Haffari, Monash University

Demonstration Chairs

Lingpeng Kong, The University of Hong Kong
Sweta Agrawal, Google
Alberto Abad, INESC-ID and Instituto Superior Técnico

Handbook Chairs

Vlad Niculae, University of Amsterdam
Jia Xu, Rutgers University

Diversity and Inclusion Chairs

Anne Lauscher, University of Hamburg
David Ifeoluwa Adelani, McGill University
Spandana Gella, ServiceNow

Student Volunteer Chairs

Nafise Sadat Moosavi, University of Sheffield
Favour Aghaebe, University of Sheffield

Website Chairs

Raj Dabre, Google
Hour Kaing, NICT

Virtual Program Chairs

Fernando Batista, Iscte and INESC-ID
Ali Hürriyetoğlu, Wageningen Food Safety Research

Sponsorship Chairs

Nikos Aletras, University of Sheffield
Matthias Gallé, Cohere

Workflow Chairs

Cheng-Han Chiang, National Taiwan University
Yating Wu, UT Austin

Publicity Chairs

Yuki Arase, Institute of Science Tokyo
Eran Hirsch, Bar-Ilan University

Publication Chairs

Shaoxiong Ji, University of Turku
Vlad Niculae, University of Amsterdam
Junjie Hu, University of Wisconsin-Madison

Ethics Chairs

Anil Ramakrishna, Meta
Giuseppe Attanasio, Instituto de Telecomunicações

Program Committee

Senior Area Chairs

David Ifeoluwa Adelani, McGill University
Rodrigo Agerri, University of the Basque Country
Wasi Uddin Ahmad, NVIDIA
Alham Fikri Aji, Mohamed bin Zayed University of Artificial Intelligence
Alan Akbik, Humboldt Universität Berlin
Nader Akoury, Cornell University
Malihe Alikhani, University of Pittsburgh and Northeastern University
Emily Allaway, University of Edinburgh
Hadi Amiri, University of Massachusetts Lowell
Prithviraj Ammanabrolu, University of California, San Diego
Sophia Ananiadou, University of Manchester
Antonios Anastasopoulos, Athena Research Center and George Mason University
Jun Araki, Bosch Research
Ehsaneddin Asgari, Hamad Bin Khalifa University
Mohammad Atari, University of Massachusetts at Amherst
Tal August, University of Illinois at Urbana-Champaign
Niranjan Balasubramanian, State University of New York, Stony Brook
Dennis Giovani Balreira, Universidade Federal do Rio Grande do Sul
David Bamman, University of California Berkeley
Chitta Baral, Arizona State University
Alberto Barrón-Cedeño, Università di Bologna
Jasmijn Bastings, Google DeepMind
David Bau, Northeastern University
Beata Beigman Klebanov, Educational Testing Service
Ahmad Beirami, Fidian
Luciana Benotti, Universidad nacional de Córdoba
Jonathan Berant, Google and Tel Aviv University
Raffaella Bernardi, Free University of Bozen Bolzano
Sumit Bhatia, Adobe Systems
Archna Bhatia, Institute for Human and Machine Cognition
Jelke Bloem, University of Amsterdam
Bernd Bohnet
Danushka Bollegala, Amazon and University of Liverpool
Houda Bouamor, Carnegie Mellon University
Chris Brew, Ohio State University, Columbus and Lexis Nexis
Emanuele Bugliarello, Google DeepMind
Hendrik Buschmeier, Universität Bielefeld
Jan Buys, University of Cape Town
Yujun Cai, The University of Queensland
Giuseppe Carenini, , University of British Columbia
Asli Celikyilmaz, MSR AI Frontiers Lab
Sarath Chandar, École Polytechnique de Montréal and Mila - Quebec Artificial Intelligence Institute
Kai-Wei Chang, University of California, Los Angeles and Amazon
Tyler A. Chang, Google DeepMind
Soravit Changpinyo, Google

Snigdha Chaturvedi, Department of Computer Science, University of North Carolina at Chapel Hill
Wenhu Chen, University of Waterloo
Hanjie Chen, Rice University
Chung-Chi Chen, NII, Tokyo Institute of Technology
Mingda Chen, Shanghai Jiao Tong University
Jackie Chi Kit Cheung, McGill University and Mila Research Institute
Eleanor Chodroff, University of Zurich
Monojit Choudhury, Mohamed bin Zayed University of Artificial Intelligence
Christos Christodoulopoulos, Information Commissioner's Office
Arman Cohan, Yale University and Allen Institute for Artificial Intelligence
Shay B. Cohen, University of Edinburgh
Cagri Coltekin, University of Tuebingen
Rolando Coto-Solano, Dartmouth College
Heriberto Cuayahuitl, University of Lincoln
Yiming Cui
Raj Dabre, Department of Computer Science, Indian Institute of Technology, Madras, Indian Institute of Technology, Madras and National Institute of Information and Communications Technology (NICT), National Institute of Advanced Industrial Science and Technology
Sandipan Dandapat, Microsoft
Sourish Dasgupta, Dhirubhai Ambani University
Shumin Deng, Zhejiang University
Tejaswini Deoskar, Utrecht University
Leon Derczynski, NVIDIA and IT University of Copenhagen
Giorgio Di Nunzio, University of Padua
Yuxiao Dong, Tsinghua University
Yue Dong, University of California, Riverside
Li Dong, Microsoft Research
A. Seza Doğruöz, Ghent University
Mengnan Du, The Chinese University of Hong Kong, Shenzhen
Kevin Duh, Johns Hopkins University
Nadir Durrani, Qatar Computing Research Institute
Nouha Dziri, Cohere
Steffen Eger, University of Technology Nuremberg
Carsten Eickhoff, Eberhard-Karls-Universität Tübingen
Asif Ekbal, Indian Institute of Technology, Jodhpur
Michael Elhadad, Ben Gurion University of the Negev
Desmond Elliott, University of Copenhagen
Micha Elsner, Ohio State University
Ali Emami, Emory University
Tianqing Fang, Tencent AI Lab
Liu Fei
Hao Fei, National University of Singapore and University of Oxford
Anna Feldman, Montclair State University
Yansong Feng, Peking University
Fuli Feng, University of Science and Technology of China
Yang Feng, Institute of Computing Technology, Chinese Academy of Sciences
Raquel Fernández, European Commission and University of Amsterdam
Francis Ferraro, University of Maryland, Baltimore County
Anjalie Field, Johns Hopkins University
Michael Flor, Educational Testing Service

Antske Fokkens, VU University Amsterdam
Markus Freitag, Google
Daniel Fried, Meta AI and Carnegie Mellon University
Zhe Gan, Apple
Kallirroi Georgila, University of Southern California and USC/ISI
Mor Geva, Tel Aviv University
Marjan Ghazvininejad, Facebook
Yoav Goldberg, Bar-Ilan University and Allen Institute for Artificial Intelligence
Dan Goldwasser, Purdue University and Purdue University
Pawan Goyal, Indian Institute of Technology Kharagpur
Kartik Goyal, Georgia Institute of Technology
Yvette Graham
Qipeng Guo, Shanghai AI Laboratory
Jiafeng Guo, Institute of Computing Technolgy, Chinese Academy of Sciences
Nizar Habash, New York University Abu Dhabi
Gholamreza Haffari, Monash University
Devamanyu Hazarika, Facebook
John Hewitt, Columbia University and Google DeepMind
Ryuichiro Higashinaka, Nagoya University and NTT Inc.
Mohammad Javad Hosseini, Google
Yufang Hou, IT:U Interdisciplinary Transformation University Austria, Technische Universität Darmstadt and IBM Research Ireland
Junjie Hu, University of Wisconsin, Madison
Jennifer Hu, Johns Hopkins University
Baotian Hu, Harbin Institute of Technology
Minlie Huang
Yinya Huang, ETH Zurich
Kuan-Hao Huang, Texas A&M University
Ruihong Huang, Texas A&M University - College Station, Department of Computer Science and Engineering, Texas A&M University - College Station
Dieuwke Hupkes, Facebook
Naoya Inoue, Ichikara, RIKEN and Japan Advanced Institute of Science and Technology
Mohit Iyyer, University of Maryland, College Park
Yangfeng Ji, University of Virginia, Charlottesville
Zhijing Jin, EuroSafeAI and University of Toronto & Vector Institute
Shafiq Joty, Nanyang Technological University and SalesForce.com
David Jurgens, University of Michigan - Ann Arbor
Preethi Jyothi, Indian Institute of Technology Bombay
Rabeeh Karimi mahabadi, Facebook
Marzena Karpinska, Simon Fraser University
Gjergji Kasneci, Technische Universität München and University of Tuebingen
Daisuke Kawahara, Waseda University
Salam Khalifa, New York University and State University of New York, Stony Brook
Omar Khattab, Massachusetts Institute of Technology
Ashiqur R. KhudaBukhsh, Rochester Institute of Technology
Joo-Kyung Kim, Amazon
Najoung Kim, Boston University
Svetlana Kiritchenko
Roman Klinger, Otto-Friedrich Universität Bamberg
Lingpeng Kong, Department of Computer Science, The University of Hong Kong
Parisa Kordjamshidi, Michigan State University

Kalpesh Krishna, Google DeepMind
Sachin Kumar, Ohio State University, Columbus
Philippe Laban, Microsoft
Matthieu Labeau, ISIR / Sorbonne Université and Télécom Paris
Jey Han Lau, The University of Melbourne
Anne Lauscher, Universität Hamburg
Carolin Lawrence, NEC Laboratories Europe and NEC Laboratories Europe
Roy Ka-Wei Lee, University of British Columbia
Hung-yi Lee, National Taiwan University
Lei Li, School of Computer Science, Carnegie Mellon University
Manling Li, Northwestern University
Chunyuan Li, Microsoft Research
Zhuang Li, Royal Melbourne Institute of Technology
Dongxu Li, Apple
Liang Li
Jing Li, The Hong Kong Polytechnic University
Constantine Lignos, Brandeis University
Bill Yuchen Lin, xAI and University of Washington
Chenghua Lin, University of Manchester
Zhouhan Lin, Shanghai Jiao Tong University
Zoey Liu, University of Florida
Zhiyuan Liu, Tsinghua University
Xuebo Liu, Harbin Institute of Technology, Shenzhen
Peng Liu, Norwegian University of Science and Technology
Qianying Liu, NII
Fangyu Liu, Google DeepMind
Kang Liu, Institute of automation, Chinese academy of science, Chinese Academy of Sciences
Yiqun Liu, Tsinghua University
Pan Lu, Stanford University
Wei Lu, Nanyang Technological University, Singapore
Ziyang Luo, Amazon
Jing Ma, Hong Kong Baptist University
Pranava Madhyastha, Alan Turing Institute, City, University of London and Imperial College London
Kyle Mahowald, The University of Texas at Austin
Lara J. Martin, University of Maryland, Baltimore County
Bruno Martins, Instituto Superior Técnico
Ninareh Mehrabi, Facebook
Tristan Miller, University of Manitoba
Sewon Min, University of California, Berkeley and Allen Institute for Artificial Intelligence
Dipendra Misra, Mosaic Research, Databricks
Kanishka Misra, University of Texas at Austin
Teruko Mitamura, Carnegie Mellon University
Tanu Mitra, University of Washington
Makoto Miwa, Toyota Technological Institute and AIST, National Institute of Advanced Industrial Science and Technology
Saif M. Mohammad
Nafise Sadat Moosavi, University of Sheffield
Alessandro Moschitti, Amazon
Aaron Mueller, Boston University
Aakanksha Naik, Allen Institute for Artificial Intelligence and National Institutes of Health

Hideki Nakayama, The University of Tokyo
Preslav Nakov, Mohamed bin Zayed University of Artificial Intelligence
Akshay Nambi, Microsoft Research
Usman Naseem, Macquarie University
Tristan Naumann, Microsoft Research
Graham Neubig, Carnegie Mellon University
Vincent Ng
Xuefei Ning, Tsinghua University
Debora Nozza, Bocconi University
Brendan O'Connor, University of Massachusetts, Amherst
Maciej Ogrodniczuk, Institute of Computer Science Polish Academy of Sciences
Naoaki Okazaki, Institute of Science Tokyo
Nedjma Ousidhoum, Cardiff University
Alexis Palmer, University of Colorado at Boulder
Alexandros Papangelis, Apple
Junhyung Park
Chan Young Park, University of Texas at Austin
Md Rizwan Parvez, Qatar Computing Research Institute
Rebecca J. Passonneau, Pennsylvania State University
Viviana Patti, Computer Science Department
Nanyun Peng, Google and University of California, Los Angeles
Hao Peng, University of Illinois Urbana-Champaign
Gabriele Pergola, University of Warwick
Sandro Pezzelle, University of Amsterdam
Mohammad Taher Pilehvar, Cardiff University, TeIAS and University of Cambridge
Tiago Pimentel, Department of Computer Science, ETHZ - ETH Zurich
Yuval Pinter, Ben-Gurion University of the Negev
Flor Miriam Plaza-del-Arco, Leiden University
Massimo Poesio, Utrecht University and Queen Mary, University of London
Soujanya Poria, Nanyang Technological University
Emily Prud'hommeaux, Boston College
Danish Pruthi, Indian Institute of Science, Bangalore
Peng Qi, Uniphore
Tao Qi, Beijing University of Posts and Telecommunications
Lianhui Qin
Libo Qin, Harbin Institute of Technology, Shenzhen
Xipeng Qiu, Shanghai Innovation Institution and Fudan University
Ariadna Quattoni, Universidad Politécnica de Cataluna
Leonardo Ranaldi
Sudha Rao, Microsoft
Abhinav Rastogi, Google
Marek Rei, Imperial College London
Roi Reichart, Technion, Israel Institute of Technology
Mehdi Rezagholizadeh, Advanced Micro Devices
Verena Rieser, Google DeepMind
Alla Rozovskaya
Frank Rudzicz, Dalhousie University
Mrinmaya Sachan, Swiss Federal Institute of Technology
Keisuke Sakaguchi, Tohoku University
Sakriani Sakti, Nara Institute of Science and Technology, Japan, Japan Advanced Institute of Science and Technology, Tokyo Institute of Technology, Universitas Indonesia and RIKEN

Naomi Saphra, Boston University, Boston University
Asad B. Sayeed, University of Gothenburg
Michael Sejr Schlichtkrull, Facebook
Natalie Schluter, Apple and IT University
Sebastian Schuster, University of Vienna
Roy Schwartz, , Hebrew University of Jerusalem
Ehsan Shareghi, University College London
Jiaming Shen, Google DeepMind
Avi Sil, Oracle
Carina Silberer
Mukul Singh, Microsoft
Koustuv Sinha, Meta (FAIR)
Gabriel Skantze, KTH Royal Institute of Technology, Stockholm, Sweden
Kevin Small, Microsoft
Yangqiu Song, Hong Kong University of Science and Technology
Richard Sproat, Sakana AI
Elias Stengel-Eskin, University of Texas at Austin
Jimeng Sun, University of Illinois, Urbana Champaign, College of Computing and Georgia Institute of Technology
Huan Sun, The Ohio State University, Columbus
Chenhao Tan, University of Chicago
Hao Tan, Adobe Systems
Buzhou Tang, Harbin Institute of Technology
Harish Tayyar Madabushi, University of Bath
Joel R. Tetreault
Sara Tonelli
Marcos Vinicius Treviso, Instituto Superior Técnico, University of Lisbon
Kewei Tu, ShanghaiTech University
Zhaopeng Tu, Tencent
Adaku Uchendu, MIT Lincoln Laboratory
Daniel Varab, Corti
Aline Villavicencio, University of Exeter and University of Sheffield
Thang Vu, University of Stuttgart, University of Stuttgart
Henning Wachsmuth, Leibniz Universität Hannover
Byron C Wallace, Northeastern University
Hao Wang, Sichuan University
Qifan Wang, Meta AI
Rui Wang, Shanghai Jiao Tong University
Xuan Wang, Virginia Polytechnic Institute and State University
Yiwei Wang, University of California, Merced
Xin Eric Wang, University of California, Santa Barbara and Simular
Benyou Wang, The Chinese University of Hong Kong, Shenzhen
Lilian Diana Awuor Wanzare, Maseno University
Alex Warstadt, University of California, San Diego
Sarah Wiegreffe, University of Maryland, College Park
Adina Williams, FAIR (Meta Platforms Inc.)
Derek F. Wong, University of Macau
Tongshuang Wu, School of Computer Science, Carnegie Mellon University
Rui Xia, Nanjing University
Chaowei Xiao, Johns Hopkins University and NVIDIA
Tong Xiao, Northeastern University

Qianqian Xie, Wuhan University
Frank F. Xu, Microsoft
Wei Xu, Georgia Institute of Technology
Nianwen Xue, Brandeis University
Diyi Yang, Stanford University
Helen Yannakoudakis, King's College London, University of London
Huaxiu Yao, Department of Computer Science, University of North Carolina at Chapel Hill
Andrew Yates, Johns Hopkins University
Wenpeng Yin, Pennsylvania State University
Dani Yogatama, Google DeepMind and DeepMind
Koichiro Yoshino, Osaka University and RIKEN
Zhou Yu, Columbia University
Kai Yu, Shanghai Jiao Tong University
Fabio Massimo Zanzotto, University of Rome Tor Vergata
Amir Zeldes, Georgetown University
Chrysoula Zerva, National Technical University of Athens
Luke Zettlemoyer, University of Washington, Facebook and Meta
Lemei Zhang, Norwegian University of Science and Technology
Yue Zhang, Westlake University
Zhuosheng Zhang, Shanghai Jiao Tong University
Lingming Zhang, University of Illinois Urbana-Champaign
Yu Zhang, Texas A&M University - College Station
Wei Emma Zhang, Adelaide University
Jiajun Zhang, Institute of automation, Chinese academy of science, Chinese Academy of Sciences
Qi Zhang, Fudan University
Weinan Zhang, Harbin Institute of Technology
Mengjie Zhao, Northeastern University and SB Intuitions / Softbank
Tuo Zhao, Georgia Institute of Technology
Xin Zhao, Renmin University of China
Zilong Zheng, Beijing Institute for General Artificial Intelligence
Yi Zhou, Cardiff University
Terry Yue Zhuo
Miryam de Lhoneux, KU Leuven

Area Chairs

David Ifeoluwa Adelani, McGill University
Wasi Uddin Ahmad, NVIDIA
Malihe Alikhani, University of Pittsburgh and Northeastern University
Emily Allaway, University of Edinburgh
Hadi Amiri, University of Massachusetts Lowell
Sophia Ananiadou, University of Manchester
Antonios Anastasopoulos, Athena Research Center and George Mason University
Jun Araki, Bosch Research
Ehsaneddin Asgari, Hamad Bin Khalifa University
Elliott Ash, Swiss Federal Institute of Technology
Tal August, University of Illinois at Urbana-Champaign
Niranjan Balasubramanian, State University of New York, Stony Brook
Chitta Baral, Arizona State University
Jasmijn Bastings, Google DeepMind
Beata Beigman Klebanov, Educational Testing Service

Ahmad Beirami, Fidian
Luciana Benotti, Universidad nacional de Córdoba
Jonathan Berant, Google and Tel Aviv University
Archna Bhatia, Institute for Human and Machine Cognition
Sumit Bhatia, Adobe Systems
Jelke Bloem, University of Amsterdam
Bernd Bohnet
Danushka Bollegala, Amazon and University of Liverpool
Houda Bouamor, Carnegie Mellon University
Chris Brew, Ohio State University, Columbus and Lexis Nexis
Hendrik Buschmeier, Universität Bielefeld
Jan Buys, University of Cape Town
Yujun Cai, The University of Queensland
Giuseppe Carenini, , University of British Columbia
Samuel Carton, University of New Hampshire, Durham
Sarath Chandar, École Polytechnique de Montréal and Mila - Quebec Artificial Intelligence Institute
Angel X Chang, Simon Fraser University
Tyler A. Chang, Google DeepMind
Kai-Wei Chang, University of California, Los Angeles and Amazon
Soravit Changpinyo, Google
Snigdha Chaturvedi, Department of Computer Science, University of North Carolina at Chapel Hill
Mingda Chen, Shanghai Jiao Tong University
Chung-Chi Chen, NII, Tokyo Institute of Technology
Hanjie Chen, Rice University
Yulong Chen, University of Aberdeen and University of Cambridge
Jackie Chi Kit Cheung, McGill University and Mila Research Institute
Cheng-Han Chiang, ByteDance Inc.
Jinho D. Choi, Emory University
Monojit Choudhury, Mohamed bin Zayed University of Artificial Intelligence
Arman Cohan, Yale University and Allen Institute for Artificial Intelligence
Shay B. Cohen, University of Edinburgh
Cagri Coltekin, University of Tuebingen
Ryan Cotterell, Swiss Federal Institute of Technology
Heriberto Cuayahuitl, University of Lincoln
Yiming Cui
Sandipan Dandapat, Microsoft
Sourish Dasgupta, Dhirubhai Ambani University
Shumin Deng, Zhejiang University
Tejaswini Deoskar, Utrecht University
Li Dong, Microsoft Research
Yue Dong, University of California, Riverside
Yuxiao Dong, Tsinghua University
A. Seza Doğruöz, Ghent University
Mengnan Du, The Chinese University of Hong Kong, Shenzhen
Kevin Duh, Johns Hopkins University
Nadir Durrani, Qatar Computing Research Institute
Carsten Eickhoff, Eberhard-Karls-Universität Tübingen
Asif Ekbal, Indian Institute of Technology, Jodhpur
Michael Elhadad, Ben Gurion University of the Negev

Micha Elsner, Ohio State University
Ali Emami, Emory University
Tianqing Fang, Tencent AI Lab
Hao Fei, National University of Singapore and University of Oxford
Anna Feldman, Montclair State University
Yang Feng, Institute of Computing Technology, Chinese Academy of Sciences
Fuli Feng, University of Science and Technology of China
Yansong Feng, Peking University
Raquel Fernández, European Commission and University of Amsterdam
Javier Ferrando, Amazon
Francis Ferraro, University of Maryland, Baltimore County
Michael Flor, Educational Testing Service
Markus Freitag, Google
Zhe Gan, Apple
Kallirroi Georgila, University of Southern California and USC/ISI
Mor Geva, Tel Aviv University
Goran Glavaš, Julius-Maximilians-Universität Würzburg
Dan Goldwasser, Purdue University and Purdue University
Kartik Goyal, Georgia Institute of Technology
Pawan Goyal, Indian Institute of Technology Kharagpur
Vivek Gupta, Arizona State University
Nizar Habash, New York University Abu Dhabi
David Harwath, University of Texas, Austin
Devamanyu Hazarika, Facebook
Ryuichiro Higashinaka, Nagoya University and NTT Inc.
Yufang Hou, IT:U Interdisciplinary Transformation University Austria, Technische Universität Darmstadt and IBM Research Ireland
Cho-Jui Hsieh, LMArena and University of California, Los Angeles
Junjie Hu, University of Wisconsin, Madison
Baotian Hu, Harbin Institute of Technology
Ruihong Huang, Texas A&M University - College Station, Department of Computer Science and Engineering, Texas A&M University - College Station
Yinya Huang, ETH Zurich
Kuan-Hao Huang, Texas A&M University
Dieuwke Hupkes, Facebook
Naoya Inoue, Ichikara, RIKEN and Japan Advanced Institute of Science and Technology
Yangfeng Ji, University of Virginia, Charlottesville
Robin Jia, University of Southern California
Zhijing Jin, EuroSafeAI and University of Toronto & Vector Institute
Preethi Jyothi, Indian Institute of Technology Bombay
Rabeeh Karimi mahabadi, Facebook
Marzena Karpinska, Simon Fraser University
Gjergji Kasneci, Technische Universität München and University of Tuebingen
Daisuke Kawahara, Waseda University
Salam Khalifa, New York University and State University of New York, Stony Brook
Omar Khattab, Massachusetts Institute of Technology
Ashiqur R. KhudaBukhsh, Rochester Institute of Technology
Joo-Kyung Kim, Amazon
Roman Klinger, Otto-Friedrich Universität Bamberg
Lingpeng Kong, Department of Computer Science, The University of Hong Kong
Parisa Kordjamshidi, Michigan State University

Kalpesh Krishna, Google DeepMind
Sachin Kumar, Ohio State University, Columbus
Philippe Laban, Microsoft
Matthieu Labeau, ISIR / Sorbonne Université and Télécom Paris
Jey Han Lau, The University of Melbourne
Carolin Lawrence, NEC Laboratories Europe and NEC Laboratories Europe
Roy Ka-Wei Lee, University of British Columbia
Hung-yi Lee, National Taiwan University
Wenqiang Lei, Sichuan University
Piyawat Lertvittayakumjorn, Google
Jing Li, The Hong Kong Polytechnic University
Lei Li, School of Computer Science, Carnegie Mellon University
Dongxu Li, Apple
Liang Li
Zhuang Li, Royal Melbourne Institute of Technology
Constantine Lignos, Brandeis University
Peng Liu, Norwegian University of Science and Technology
Xuebo Liu, Harbin Institute of Technology, Shenzhen
Zhiyuan Liu, Tsinghua University
Fangyu Liu, Google DeepMind
Fei Liu, Emory University
Yiqun Liu, Tsinghua University
Kang Liu, Institute of automation, Chinese academy of science, Chinese Academy of Sciences
Zoey Liu, University of Florida
Wei Lu, Nanyang Technological University, Singapore
Pan Lu, Stanford University
Ziyang Luo, Amazon
Jing Ma, Hong Kong Baptist University
Joel Mackenzie, The University of Queensland
Pranava Madhyastha, Alan Turing Institute, City, University of London and Imperial College London
Kyle Mahowald, The University of Texas at Austin
Bruno Martins, Instituto Superior Técnico
Yuval Marton, University of Washington
Sérgio Matos, Universidade de Aveiro
Ninareh Mehrabi, Facebook
Florian Metze, Carnegie Mellon University and Meta Platforms
Tristan Miller, University of Manitoba
Sewon Min, University of California, Berkeley and Allen Institute for Artificial Intelligence
Pasquale Minervini, University of Edinburgh, University of Edinburgh
Kanishka Misra, University of Texas at Austin
Dipendra Misra, Mosaic Research, Databricks
Teruko Mitamura, Carnegie Mellon University
Tanu Mitra, University of Washington
Makoto Miwa, Toyota Technological Institute and AIST, National Institute of Advanced Industrial Science and Technology
Marie-Francine Moens, KU Leuven, KU Leuven
Saif M. Mohammad
Nafise Sadat Moosavi, University of Sheffield
Marius Mosbach, McGill University and Mila - Quebec Artificial Intelligence Institute
Alessandro Moschitti, Amazon

Aaron Mueller, Boston University
Aakanksha Naik, Allen Institute for Artificial Intelligence and National Institutes of Health
Hideki Nakayama, The University of Tokyo
Preslav Nakov, Mohamed bin Zayed University of Artificial Intelligence
Akshay Nambi, Microsoft Research
Usman Naseem, Macquarie University
Tristan Naumann, Microsoft Research
Vincent Ng
Xuefei Ning, Tsinghua University
Debora Nozza, Bocconi University
Maciej Ogrodniczuk, Institute of Computer Science Polish Academy of Sciences
Alice Oh, Google and Korea Advanced Institute of Science and Technology
Naoaki Okazaki, Institute of Science Tokyo
Manabu Okumura, Institute of Science Tokyo and Tokyo Institute of Technology, Tokyo Institute of Technology
Nedjma Ousidhoum, Cardiff University
Sebastian Padó, University of Stuttgart, Universität Stuttgart
Alexis Palmer, University of Colorado at Boulder
Alexandros Papangelis, Apple
Chan Young Park, University of Texas at Austin
Md Rizwan Parvez, Qatar Computing Research Institute
Viviana Patti, Computer Science Department
Nanyun Peng, Google and University of California, Los Angeles
Hao Peng, University of Illinois Urbana-Champaign
Gabriele Pergola, University of Warwick
Sandro Pezzelle, University of Amsterdam
Mohammad Taher Pilehvar, Cardiff University, TeIAS and University of Cambridge
Tiago Pimentel, Department of Computer Science, ETHZ - ETH Zurich
Yuval Pinter, Ben-Gurion University of the Negev
Massimo Poesio, Utrecht University and Queen Mary, University of London
Soujanya Poria, Nanyang Technological University
Emily Prud'hommeaux, Boston College
Danish Pruthi, Indian Institute of Science, Bangalore
Tao Qi, Beijing University of Posts and Telecommunications
Peng Qi, Uniphore
Lianhui Qin
Libo Qin, Harbin Institute of Technology, Shenzhen
Xipeng Qiu, Shanghai Innovation Institution and Fudan University
Ariadna Quattoni, Universidad Politécnica de Cataluna
Leonardo Ranaldi
Hannah Rashkin, Google Deepmind
Abhinav Rastogi, Google
Shauli Ravfogel, New York University
Marek Rei, Imperial College London
Roi Reichart, Technion, Israel Institute of Technology
Zhaochun Ren, Leiden University
Mehdi Rezagholizadeh, Advanced Micro Devices
Alla Rozovskaya
Frank Rudzicz, Dalhousie University
Mrinmaya Sachan, Swiss Federal Institute of Technology
Hassan Sajjad, Dalhousie University

Sakriani Sakti, Nara Institute of Science and Technology, Japan, Japan Advanced Institute of Science and Technology, Tokyo Institute of Technology, Universitas Indonesia and RIKEN
Naomi Saphra, Boston University, Boston University
Sebastian Schuster, University of Vienna
Roy Schwartz, , Hebrew University of Jerusalem
Rico Sennrich, University of Zurich
Ehsan Shareghi, University College London
Jiaming Shen, Google DeepMind
Carina Silberer
Mukul Singh, Microsoft
Koustuv Sinha, Meta (FAIR)
Sunayana Sitaram, Microsoft
Gabriel Skantze, KTH Royal Institute of Technology, Stockholm, Sweden
Kevin Small, Microsoft
Yangqiu Song, Hong Kong University of Science and Technology
Elias Stengel-Eskin, University of Texas at Austin
Emma Strubell, Carnegie Mellon University
Jimeng Sun, University of Illinois, Urbana Champaign, College of Computing and Georgia Institute of Technology
Fei Sun, Institute of Computing Technology, Chinese Academy of Sciences
Huan Sun, The Ohio State University, Columbus
Buzhou Tang, Harbin Institute of Technology
Harish Tayyar Madabushi, University of Bath
Joel R. Tetreault
Sara Tonelli
Marcos Vinicius Treviso, Instituto Superior Técnico, University of Lisbon
Zhaopeng Tu, Tencent
Kewei Tu, ShanghaiTech University
Adaku Uchendu, MIT Lincoln Laboratory
Daniel Varab, Corti
Thang Vu, University of Stuttgart, University of Stuttgart
Henning Wachsmuth, Leibniz Universität Hannover
Byron C Wallace, Northeastern University
Hao Wang, Sichuan University
Lucy Lu Wang, University of Washington and Allen Institute for Artificial Intelligence
Xuan Wang, Virginia Polytechnic Institute and State University
Yiwei Wang, University of California, Merced
Qifan Wang, Meta AI
Xin Eric Wang, University of California, Santa Barbara and Simular
Rui Wang, Shanghai Jiao Tong University
Benyou Wang, The Chinese University of Hong Kong, Shenzhen
Alex Warstadt, University of California, San Diego
Sarah Wiegreffe, University of Maryland, College Park
Derek F. Wong, University of Macau
Tongshuang Wu, School of Computer Science, Carnegie Mellon University
Rui Xia, Nanjing University
Chaowei Xiao, Johns Hopkins University and NVIDIA
Tong Xiao, Northeastern University
Qianqian Xie, Wuhan University
Frank F. Xu, Microsoft
Wei Xu, Georgia Institute of Technology

Nianwen Xue, Brandeis University
Rui Yan, Wuhan University
Min Yang, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Chinese Academy of Sciences
Diyi Yang, Stanford University
Huaxiu Yao, Department of Computer Science, University of North Carolina at Chapel Hill
Andrew Yates, Johns Hopkins University
Wenpeng Yin, Pennsylvania State University
Fabio Massimo Zanzotto, University of Rome Tor Vergata
Amir Zeldes, Georgetown University
Chrysoula Zerva, National Technical University of Athens
Luke Zettlemoyer, University of Washington, Facebook and Meta
Lingming Zhang, University of Illinois Urbana-Champaign
Yu Zhang, Texas A&M University - College Station
Zhuosheng Zhang, Shanghai Jiao Tong University
Weinan Zhang, Harbin Institute of Technology
Qi Zhang, Fudan University
Jiajun Zhang, Institute of automation, Chinese academy of science, Chinese Academy of Sciences
Yue Zhang, Westlake University
Wei Emma Zhang, Adelaide University
Lemei Zhang, Norwegian University of Science and Technology
Xin Zhao, Renmin University of China
Mengjie Zhao, Northeastern University and SB Intuitions / Softbank
Zilong Zheng, Beijing Institute for General Artificial Intelligence
Yi Zhou, Cardiff University
Zining Zhu, Stevens Institute of Technology
Terry Yue Zhuo
Miryam de Lhoneux, KU Leuven

Keynote 1

Towards AI that understands the Real World

Pascale Fung
AMI Labs and The Hong Kong University of Science and Technology



Sun, October 25, 2026 – Time: **09:30 – 10:30** –

Abstract: Recent advances in AI have been driven by large-scale generative models with impressive capabilities in language, code, and digital tasks. Yet these systems still have a limited understanding of the physical world, human intentions, and real-world context. In this talk, Pascale Fung will discuss the next frontier of AI: world models. World models allow AI systems to learn abstract representations of the world from continuous multimodal data, enabling reasoning, planning, memory, and action beyond next-token prediction. Moving toward real-world AI agents requires modeling not only the physical world, but also the mental world—including intent, attention, goals, and social context. The future of AI will depend on systems that can understand, adapt, and collaborate in the real world, safely and reliably.

Bio: Pascale Fung's long term research background is in multimodal interactive systems including audio, speech, text and video. She started research on world modeling after studying the limitation of generative models due to hallucinations. She is the Co-founder and Chief Research & Innovation Officer at AMI Labs. She was previously the Senior Director of AI Research at Meta-FAIR, leading research on embodied AI agents. She is also a Chair Professor of ECE at The Hong Kong University of Science & Technology (HKUST). She is a Fellow of the ACL, AAAI, IEEE, and ISCA for her significant contribution to human-machine interactions.

Keynote 2

On the Value of Open Research in Language Modeling - An Adversarial Dialog

Graham Neubig
Carnegie Mellon University and OpenHands



Mon, October 26, 2026 – Time: **15:15 – 16:15** –

Abstract: Large language models are simultaneously one of the greatest advances in natural language processing ever and one of the most confounding for the open research community. Given the extreme resource intensiveness and economic value of research in this area, much research has moved out of the public eye and behind closed doors. What does that mean for scientists, like many of the participants in EMNLP, who believe in the value of open publishing of research results? In this talk, I will present some often-cited arguments against the value of open research in core LLM technology, and some counter arguments for why open research is still important, albeit under certain conditions. I will also try to present some personal anecdotes for my own projects that I think have been successful in the face of these headwinds, and may be informative for listeners who are also interested in expanding the impact of their work.

Bio: Graham Neubig is an associate professor at the Language Technologies Institute of Carnegie Mellon University and Chief Scientist at OpenHands. His research focuses on large language models, including both fundamental advances in model capabilities and applications to tasks such as software development. His final goal is that every person in the world should be able to communicate with each other, and with computers in their own language. He also contributes to making NLP research more accessible through open publishing of research papers, advanced NLP course materials and video lectures, and open-source software, all of which are available on his web site.

Keynote 3

AI for the Many: Starting from the World, Not the Model

Anna Korhonen
University of Cambridge



Tue, October 27, 2026 – Time: **14:30 – 15:30** –

Abstract: AI is advancing at extraordinary speed, but its benefits remain profoundly uneven. For much of the world, the dominant path towards ever larger, more resource-intensive models is out of reach and not necessarily the right one. In this talk, I will examine the global AI divide through the lens of language. Recent evidence shows a widening global linguistic hierarchy, with thousands of languages and their communities at risk of being left behind. I will argue that addressing this divide requires a fundamental shift in how we approach AI development. Instead of asking how existing models can be extended to more people, we should ask what communities need from AI and how it can deliver meaningful benefit in their particular contexts. This means grounding AI in local needs and realities, taking account of available resources and infrastructure, and judging success by what it delivers in practice. If this technology is to make a meaningful difference across the world, we need not one path to AI, but many.

Bio: Anna Korhonen is Professor of Natural Language Processing at the University of Cambridge. Her research focuses on multilingual and human-centred language technology and AI, with the broader aim of ensuring societal and global benefit. Her work spans applications in health, education, science and the environment. She is the Director and co-founder of the Centre for Human-Inspired Artificial Intelligence (CHIA), the Institute for Technology and Humanity (ITH), and the Cambridge Language Technology Laboratory (LTL). She is a Fellow of the ACL, ELLIS and the British Academy. Her work informs global AI policy and governance through the United Nations and other international bodies.

Panel

New Missions in NLP Research

Sun, October 25, 2026 – Time: **14:30 – 15:30** – Room: **Hall H**

Large language models have rapidly shifted from research prototypes to widely used infrastructure. This new reality is changing the center of gravity of NLP research, and progress can no longer be measured only by fractional improvements in benchmark scores. When strong general-purpose models are increasingly available through open releases and commercial platforms, the field faces a more foundational question: What are the missions of NLP research now, and what kinds of contributions most advance that mission?
This panel will focus on EMNLP 2026's special theme to discuss and advance visions for NLP in this era.

Moderator: Raymond J. Mooney, **The University of Texas at Austin** Raymond J. Mooney is a Professor Emeritus in the Department of Computer Science at the University of Texas at Austin. He received his Ph.D. in 1988 from the University of Illinois at Urbana-Champaign. He is an author of over 200 published research papers, primarily in the areas of machine learning and natural language processing. He was the President of the International Machine Learning Society from 2008–2011, program co-chair for AAAI 2006, general chair for HLT-EMNLP 2005, and co-chair for ICML 1990. He is a Fellow of AAAI, ACM, and ACL and the recipient of the Classic Paper award from AAAI-19 and best paper awards from AAAI-96, KDD-04, ICML-05 and ACL-07.

Panelists:

Dirk Hovy, Bocconi University Dirk Hovy is a Professor in the Computing Sciences Department of Bocconi University in Milan, where he co-leads the MilaNLP lab. He is also the scientific director of the Data and Marketing Insights research unit. He is interested in what computers can tell us about language and what language can tell us about society. That includes questions of bias and algorithmic fairness in machine learning. He has authored over 150 articles on these topics, including 3 best and one outstanding paper awards, and has also published two textbooks on text processing in Python for social scientists. He has co-founded and organized several workshops (on computational social science, and ethics in NLP), and was a local organizer for EMNLP 2017 and general chair for EMNLP 2025.

Lillian Lee, Cornell University Lillian Lee is the Charles Roy Davis professor of computer science at Cornell University. Her research interests include natural language processing and computational social science. She is a AAAI Fellow, an ACL Fellow, and an ACM Fellow. She is a recipient of best paper awards at ACL 2023 (joint with Jack Hessel, Ana Marasović, Jena Hwang, Jeff Da, Rowan Zellers, Bob Mankoff, and Yejin Choi, one of three awarded), NAACL 2004 (joint with Regina Barzilay, inaugural, one awarded), and the IJCAI 2016 "Natural Language Processing Meets Journalism" workshop (joint with Liye Fu and Cristian Danescu-Niculescu-Mizil, one awarded); Test of Time awards from NAACL 2018 for a paper between 2002–2012 (joint with Bo Pang, inaugural, one of three awarded) and ACL 2024 for 25 years (one awarded); and the ACL 2021 Distinguished Service Award for her work with TACL. Her co-authored work has received several mentions in the popular press, including The New York Times, NPR's All Things Considered, and NBC's The Today Show; and one of her co-authored papers on the memorability of movie quotes was publicly called "boring" by YouTube channel Good Mythical Morning in a video viewed over 2.7 million times.

Mor Geva, Tel Aviv University Mor Geva is an Assistant Professor at the School of Computer Science and AI at Tel Aviv University. Her research focuses on understanding the inner workings of large lan-

guage models to increase their transparency, control their behavior, and improve their reasoning abilities. Mor completed a Ph.D. in Computer Science at Tel Aviv University, was a postdoctoral researcher at Google DeepMind and the Allen Institute for AI, and worked as a Research Scientist at Google Research. She is a recipient of Intel's Rising Star Faculty Award (2024), the Alon Scholarship for Outstanding Faculty (2024), EMNLP Best Paper Award (2024), EACL Outstanding Paper Award (2023), MIT Rising Star in EECS nomination (2021), and the Dan David Prize for Graduate Students in the field of AI (2020).

Table of Contents

DeepSEA: A Multilingual Evaluation Benchmark for Southeast Asia
Jiuzhou Han, Abhishek Bijay Mishra, Trevor Cohn and Shubham Mittal 1

PopResume: Causal Fairness Evaluation of LLM/VLM Resume Screeners with Population-Representative Dataset
Sumin Yu, Juhyeon Park and Taesup Moon 18

Logic Meets Language: Hybrid Query Answering over Incomplete Text-Labeled Graphs
Yunjie He, Bo Xiong, Daniel Hernández, Yuqicheng Zhu, Yi Wang, Evgeny Kharlamov and Steffen Staab 40

From Self to Other: Evaluating Demographic Perspective-Taking in LLM Hate Speech Annotation
Paloma Piot and Javier Parapar 64

The Hidden Toll of Social Media News on Psychosocial Wellbeing
Olivia Pal, Agam Goyal, Eshwar Chandrasekharan and Koustuv Saha 81

FloCA: Towards Faithful and Logically Consistent Flowchart Reasoning
Jinzi Zou, Bolin Wang, Shuo Zhang, Nuo Xu and Junzhou Zhao 96

The Trilemma of Truth in Large Language Models
Germans Savcisens and Tina Eliassi-Rad 120

Functional Entropy: Predicting Functional Correctness in LLM-Generated Code with Uncertainty Quantification
Dylan Bouchard, Mohit Singh Chauhan, Zeya Ahmad and Ho-Kyeong Ra 195

Benchmarking Bengali Dialectal Bias: A Multi-Stage Framework Integrating RAG-Based Translation and Human-Augmented RLAIF
K. M. Jubair Sami, Dipto Sumit, Ariyan Hossain and Farig Sadeque 218

CALM: Class-Conditional Sparse Attention Vectors for Large Audio-Language Models
Videet Mehta, Liming Wang, Hilde Kuehne, Rogerio Feris, James R. Glass and Muhammad Jehanzeb Mirza 234

Check The Scoreboard: An Analysis of Scoring Schemes on Multiple-Choice Evaluation
Nishant Balepur, Paiheng Xu, Wei Ai, Eunsol Choi, Rachel Rudinger and Jordan Boyd-Graber 246

Dual-Modality Multi-Stage Adversarial Safety Training: Robustifying Multimodal Web Agents Against Cross-Modal Attacks
Haoyu Liu, Dingcheng Li, Lukas Rutishauser and Zeyu Zheng 277

Learning Personalized Prompts for Healthcare Guidance
Ruize Shi, Hong Huang, Wei Zhou, Kehan Yin, Kai Zhao and Yun Zhao 304

Expert Personas Improve LLM Alignment but Damage Accuracy: Bootstrapping Intent-Based Persona Routing with PRISM
Zizhao Hu, Mohammad Rostami and Jesse Thomason 320

Dementia-R1: Reinforced Pretraining and Reasoning from Unstructured Clinical Notes for Real-World Dementia Prognosis
Choonghan Kim, Hyunmin Hwang, Hangeol Chang, Jaemin Kim, Jinse Park, Jae-Sung Lim and Jong Chul Ye 340

Granuscore: A Reference-Free Measure of Granularity for Text Analysis and Question Answering
Lukas Ellinger, Alexander M. Fichtl, Miriam Anschütz and Georg Groh 373

Survival-Guided Length Control for Efficient Diffusion Language Models
Ivan Kobyzev, Abbas Ghaddar and Yufei Cui 391

Measuring the Depth of LLM Unlearning via Activation Patching
Jaeung Lee, Dohyun Kim and Jaemin Jo 400

Compressing Sequences in the Latent Embedding Space: K-Token Merging for Large Language Models
Zihao Xu, John Harvill, Ziwei Fan, Yizhou Sun, Hao Ding and Hao Wang 419

AI Psychosis: Does Conversational AI Amplify Delusion-Related Language?
Soorya Ram Shimgekar, Vipin Gunda, Jenny Jiwon Kim, Violeta J. Rodriguez, Hari Sundaram and Koustuv Saha 430

Dissecting Transformers: A CLEAR Perspective Towards Green AI
Hemang Jain, Shailender Goyal, Divyansh Pandey and Karthik Vaidhyanathan 452

Learning Diagnostic Clarifying Questions for Conversational Question Answering via Data-Efficient Dynamic Distillation
Jihyeong Jeon, Cheol Ryu and U Kang 484

PolySQL: Scaling Text-to-SQL Evaluation Across SQL Dialects via Automated Backend Isomorphism
Yotam Perlitz, Elad Venezian, Corentin Royer, Francesco Fusco and Andrea Giovannini 499

Improving Diffusion Language Model Decoding through Joint Search in Generation Order and Token Space
Yangyi Shen, Tianjian Feng, Jiaqi Han, Wen Wang, Tianlang Chen, Chunhua Shen, Jure Leskovec and Stefano Ermon 513

PsyAgent: Constructing Human-like Agents Based on Psychological Modeling and Contextual Interaction
Zibin Meng and Kani Chen 533

CRU-Reason: Knowledge Graph-Grounded and Interpretable Clinical Prediction with Clinical Rationale Units
Zihang Liu, Linhao Luo, Zhangchi Qiu, Shigeng Chen and Haishuai Wang 568

From Text to Structure: The PACKaGE Benchmark for Bidirectional Causal Reasoning in Paragraphs and Graphs
Juhyeon Kim, Dohun Lee, Jiwon Jang, Joseph Shim and Hyunwoo Park 585

Efficient Iterative Retrieval with Heterogeneous Batching
Dohyun Park, Hubertus Franke, Daniel G Waddington, Swaminathan Sundararaman and Yongjoo Park 612

Spexis: Speculative Lookahead Scheduling for LLM Inference
Hyungyu Jung, Jaehyeok Yu, Hoonseo Choi, Sungkyun Kim, Jinho Lee and Jiwon Seo 627

When Does Auxiliary Classification Help Regression in Practice? A Case Study in Arabic Dialectness Modeling
Abdelrahman Sakr, Marwan Torki and Nagwa El-Makky 639

Cross-Attention Eigenvalues Know the Word Order in Multilingual Machine Translation
Nikolai Prudnikov and Oleg Kachan 650

Gold vs. Translation: Are We Measuring Language, Knowledge, or Locality in Translated Benchmarks?
Klaudia-Doris Thellmann, Bernhard Stadler, Dan Saattrup Smart, Michael Färber and Jens Lehmann 657

From Passive Response to Proactive Correction: Enhancing LLM Robustness Against Input Fact Perturbations
Ping Wang, Xiangguo Sun, Bingbing Xu, Guocong Li and Xiaofeng Meng 689

How Human-Like Are Large Language Models? A Register-Aware Linguistic Evaluation Framework
Björn Nieth, Marianna Gracheva, Michaela Mahlberg, Bjoern Eskofier and Emmanuelle Salin 715

Ontology-Driven Structural Regularization for Document-Level Relation Extraction
Laura Menotti, Stefano Marchesin and Gianmaria Silvello 756

SCALE: Structured Case-law-guided Anticipatory Legal Reasoning
Mann Khatri, Santosh T.Y.S.S, Oana Ichim, Rajiv Ratn Shah and Ponnurangam Kumaraguru 774

Reasoning Under Uncertainty: Exploring Probabilistic Reasoning Capabilities of LLMs
Mobina Pournemat, Keivan Rezaei, Gaurang Sriramanan, Arman Zarei, Hamid Eghbalzadeh, Yang Wang, Jiaxiang Fu, Fuchun Peng, Zhigang Wang, Xue Feng, Nima Noorshams and Soheil Feizi 811

E^3-Pruner: Towards Efficient, Economical, and Effective Layer Pruning for Large Language Models
Tao Yuan, Haoli Bai, Pan Yinfei, Xuyang Cao, Tianyu Zhang, Lu Hou, Ting Hu and Xianzhi Yu 830

Forget What's Sensitive, Remember What Matters: Token-Level Differential Privacy in Memory Sculpting for Continual Learning
Bihao Zhan, Jie Zhou, Junsong Li, Yutao Yang, Shilian Chen, Qianjun Pan, Xin Li, Wen Wu, Xingjiao Wu, Qin Chen, Hang Yan, Liang Dou and Liang He 847

Learning to Zoom Efficiently with a Contrastive Curriculum
Falko Helm and Iryna Gurevych 862

From Leaky Thoughts to Private Reasoning: Controlling What LRMs Say to Themselves
Haritz Puerto, Haonan Li, Xudong Han, Timothy Baldwin and Iryna Gurevych 885

OpenResearcher: A Fully Open Pipeline for Long-Horizon Deep Research Trajectory Synthesis
Zhuofeng Li, Dongfu Jiang, Xueguang Ma, Haoxiang Zhang, Ping Nie, Yuyu Zhang, Kai Zou, Jianwen Xie, Yu Zhang and Wenhu Chen 908

Do Implicit Personalization and Explicit Styles Conflict? PsPLUG: A Lightweight Plug-in for Balancing Personalization and Style in Customized LLMs
Yutong Song, Jiang Wu, Shaofan Yuan, Chengze Shen, Jian Wang, Yu Wang, Nikil Dutt and Amir M. Rahmani 938

Retrieving Facts or Reinforcing Beliefs? Towards Understanding Confirmation Bias in RAG Systems for Scientific Claim Verification
Qile Wan, Davide Venditti, Fabio Massimo Zanzotto, Iryna Gurevych and Lizhen Qu 958

Client Simulation via Conversational Trajectory Modeling for Both Trainee Practice and Model Evaluation in Mental Health Counseling
Huachuan Qiu, Zhaoming Chen, Yuqian Chen, Yuan Xie, Yu Lu and Zhenzhong Lan 975

MEMORA: A Memory-Enhanced Multimodal Committee Reranking Agent for Reasoning-Intensive Retrieval
Mohamed Mahmoud, Mostafa Farouk Senussi, Mahmoud Abdalla, Mahmoud SalahEldin Kasem, Abdelrahman Abdallah and Hyun Soo Kang 1017

SpecSteer: Synergizing Local Context and Global Reasoning for Efficient Personalized Generation
Hang Lv, Sheng Liang, Hao Wang, Yongyue Zhang, Hongchao Gu, Wei Guo, Defu Lian, Yong Liu and Enhong Chen 1029

Revisiting Backdoor Attacks on LLMs: A Stealthy and Practical Poisoning Framework via Harmless Inputs
Jiawei Kong, Hao Fang, Xiaochen Yang, Kuofeng Gao, Bin Chen, Shu-Tao Xia, Yaowei Wang and Min Zhang 1047

Exposía: Teaching and Assessment of Academic Writing Skills for Research Project Proposals and Peer Feedback
Dennis Zyska, Alla Rozovskaya, Ilia Kuznetsov and Iryna Gurevych 1076

GR-SAP: Generative Replay for Safety Alignment Preservation during Fine-Tuning
Zhouxiang Fang, Jiawei Zhou and Hanjie Chen 1131

What Makes a Good Response? An Empirical Analysis of Quality in Qualitative Interviews
Jonathan Ivey, Anjalie Field and Ziang Xiao 1154

The Autonomy Tax: Defense Training Breaks LLM Agents
Li Li and Yue Zhao 1179

Multi-Expert Conformal Risk Control for Pairwise LLM Judging in Open-Ended Dialogue
Ming Cheng, Yusheng Dai, Qiuhong Ke, Zhaolin Chen and Lizhen Qu 1201

ReGround: Grounding Reviewer Comments in Multimodal Evidence
Serwar Basch, Lizhen Qu and Iryna Gurevych 1216

Capable but Careless: Do Computer-Use Agents Follow Contextual Integrity?
Anmol Goel and Iryna Gurevych 1235

PIVOTS-Bench: Evaluating Fine-Grained Interpersonal Relationship Reasoning in Multimodal Large Language Models
Shuxiang Zhang, Ciossa Yin, Wenxuan Song, Yuhang Wu and Miao Liu 1267

Retrieval-Augmented LLM Agents: Learning to Learn from Experience
Thomas Palmeira Ferraz, Romain Deffayet, Vassilina Nikoulina, Hervé Déjean and Stéphane Clinchant 1297

Conformal Privacy Auditing: Calibrated Re-identification Attacks with Statistical Guarantees
Shuo Huang, Gholamreza Haffari, Xingliang Yuan and Lizhen Qu 1336

SocialCritic: Grounding Multi-Agent Dialogue via Advantage-Based Endorsement Suppression
Shuang Cao and Rui Li 1352

Demographic Prompting at Scale: When More Attributes Hurt LLM–Human Agreement
Mahammed Kamruzzaman, Shrabon Kumar Das and Gene Louis Kim 1373

SGM: Safety Glasses for Multimodal Large Language Models via Neuron-Level Detoxification
Hongbo Wang, AprilPyone MaungMaung and Isao Echizen 1400

The Evaluation Scaling Law: Toward More Reliable LLM Assessment via Multi-Prompt Diversity
Hongzhi Luan, Zhaoxin Huan, Linfeng Shi, Longfei Zheng, Xiaolu Zhang and Jun Zhou . . . 1419

ALEE: Any-Language Evaluation of Embeddings via English-Centric Minimal Pairs
Andrianos Michail, Stylianos Psychias, Michelle Wastl, Simon Clematide, Rico Sennrich and Juri Opitz 1434

Decompose, Look, and Reason: Reinforced Latent Reasoning for VLMs
Mengdan Zhu, Senhao Cheng and Liang Zhao 1459

CoPRM: A Collaborative Process Reward Model for Guiding Search Agent Reasoning
Xingquan Zhang, Yang Zhao, Chong Li, Donglei Yu, Pu Jian and Chengqing Zong 1472

Who Laughs with Whom? Disentangling Influential Factors in Humor Preferences across User Clusters and LLMs
Soichiro Murakami, Hidetaka Kamigaito, Hiroya Takamura and Manabu Okumura 1505

ClaimFlow: Tracing the Evolution of Scientific Claims in NLP
Aniket Pramanick, Yufang Hou, Saif M. Mohammad and Iryna Gurevych 1531

T-MAP: Red-Teaming LLM Agents with Trajectory-aware Evolutionary Search
Hyomin Lee, Sangwoo Park, Yumin Choi, Sohyun An, Seanie Lee and Sung Ju Hwang 1558

Mini-BEHAVIOR-Gran: Revealing U-Shaped Effects of Instruction Granularity on Language-Guided Embodied Agents
Sukai Huang, Chenyuan Zhang, Fucai Ke, Zhixi Cai, Gholamreza Haffari, Lizhen Qu and Hamid Rezatofighi 1594

Diverging Transformer Predictions for Human Sentence Processing: A Comprehensive Analysis of Agreement Attraction Effects
Titus von der Malsburg and Sebastian Padó 1618

NanoVDR: Distilling a 2B Vision-Language Retriever into a 70M Text-Only Encoder for Visual Document Retrieval
Zhuchenyang Liu, Yao Zhang and Yu Xiao 1631

Evaluating LLM-generated Explanatory Dialogue Turns through Dialogue Completion
Yi-Sheng Hsu, Nils Feldhus and Manfred Stede 1649

Who Tests the Testers? Systematic Enumeration and Coverage Audit of LLM Agent Tool Call Safety
Xuan Chen, Lu Yan, Ruqi Zhang and Xiangyu Zhang 1680

LogicEnvGen: Task-Logic Driven Generation of Diverse Simulated Environments for Embodied AI
Jianan Wang, Siyang Zhang, Bin Li, Juan Chen, Jingtao Qi, Zhuo Zhang and Chen Qian . . . 1701

Turning the Spell Around: Lightweight Alignment Amplification via Rank-One Safety Injection
Harethah Abu Shairah, Hasan Abed Al Kader Hammoud, George Turkiyyah and Bernard Ghanem 1722

Efficient LLM Safety Evaluation through Multi-Agent Debate
Dachuan Lin, Guobin Shen, Zihao Yang, Tianrong Liu, Dongcheng Zhao and Yi Zeng 1739

TTSR: Test-Time Self-Evolving via Reflection
Haoyang He, Zihua Rong, Yunjia Zhao, Lan Yang, Jian Chang and Honggang Zhang 1756

Library Hallucinations in LLM-Generated Code: A Risk Analysis Grounded in Developer Queries
Lukas Twist, Mark Harman, Helen Yannakoudakis and Jie M. Zhang 1785

What Drives Representation Steering? A Mechanistic Case Study on Steering Refusal
Stephen S Cheng, Sarah Wiegreffe and Dinesh Manocha 1813

Calibrated Tree-Structured Speculative Decoding: Mitigating Overconfidence via Path-Level Calibration
Zixian LI, Xiaohui Song, Chi Xie, Ke Chen, Chen Chen and Haonan Lu 1844

Fast and Focused: Interactive Repository-Level Fill-in-the-Middle Code Completion with Hybrid Structural Retrieval
Long S. T. Nguyen, Dat Huynh, Quynh Vo, LE Dang Nguyen Hai, Cong-Duy T Nguyen, Anh Tuan Luu, Anh Nguyen-Duc and Tho Quan 1856

Preserving Historical Truth: Detecting Historical Revisionism in Large Language Models
Francesco Ortu, Joeun Yook, Punya Syon Pandey, Bernhard Schölkopf, Alberto Cazzaniga, Rada Mihalcea and Zhijing Jin 1875

Observable Patterns Are Not Explanations: A Causal-Geometric Analysis of Latent Reasoning Models
Darpan Aswal, Thomas Palmeira Ferraz, Yongxin Zhou and Maxime Peyrard 1898

Gated Residual Discrepancy for Efficient Bi-Encoder Factuality Verification
Zeyang Liu, Jingfeng Xue, Jiarun Fu, Wenbiao Du, Xiuqi Yang, Yifeng Fu, Junbao Chen and Yong Wang 1939

Graph2Counsel: Clinically Grounded Synthetic Counseling Dialogue Generation from Client Psychological Graphs
Aishik Mandal, Hiba Arnaout, Clarissa W. Ong, Juliet Bockhorst, Kate Sheehan, Rachael Moldow, Tanmoy Chakraborty and Iryna Gurevych 1953

LLM-Based Adversarial Persuasion Attacks on Fact-Checking Systems
João A. Leite, Olesya Razuvayevskaya, Kalina Bontcheva and Carolina Scarton 2013

SpokenUS: A Spoken User Simulator for Task-Oriented Dialogue
Jonggeun Lee, Junseong Pyo, Jeongmin Park and Yohan Jo 2035

LENS: Exploiting the Cross-View Stability Gap for Hallucination Mitigation in Medical Vision–Language Models
Dwarikanath Mahapatra, Abhijit Das, Sudipta Roy, Behzad Bozorgtabar and Imran Razzak . 2071

CORE-T: COherent REtrieval of Tables for Text-to-SQL
Hassan Soliman, Vivek Gupta, Dan Roth and Iryna Gurevych 2082

Unsupervised Speech Recognition at the Syllable Level
Liming Wang, Kai-Wei Chang, Kunio Kashino, David Harwath, Mark A. Hasegawa-Johnson and James R. Glass 2115

HybridThink: Learned Interleaving of Continuous and Discrete Reasoning in Large Language Models
Jing Gao, Weiqiang Jin, Guoshuai Zhao, Peigen Ye, Xueming Qian and Hongyang Du 2136

Entropic properties of languages in phonetic versus orthographic alphabets
Carla Olabe and Christian Bentz 2165

ViD: Vision-Dominant Gender Bias Mitigation for Large Vision-Language Models
Zhipeng Zhao, Zhaoqiang Wei, Peishun Liu, Youwei Zhao and Ruichun Tang 2184

Learning to Detect Unseen Jailbreak Attacks in Large Vision-Language Models
Shuang Liang, Zhihao Xu, Jiaqi Weng, Jialing Tao, Hui Xue and Xiting Wang 2198

Are Non-English Papers Reviewed Fairly? Language-of-Study Bias in NLP Peer Reviews
Ehsan Barkhordar, Abdulfattah Safa, Verena Blaschke, Erika Lombart, Marie-Catherine de Marneffe and Gözde Gül Şahin 2215

Self-Calibrating Language Models via Test-Time Discriminative Distillation
Mohamed Rissal Hedna, Jan Strich, Martin Semmann and Chris Biemann 2238

VITAL: Visual-Semantic Dual Supervision for Enhanced and Interpretable Latent Reasoning in Medical MLLMs
Qiaoru Li, Shaotian Liang, Jintao Chen, Haoran Sun, Yuxiang Cai, Jianwei Yin and Yankai Jiang 2267

Structure-Aware Data Augmentation for Argument Mining
Masayuki Kawarada, Tsutomu Hirao and Manabu Okumura 2295

From Reasoning to Generalization: Knowledge-Augmented LLMs for the ARC Benchmark
Chao Lei, Krista A. Ehinger, Nir Lipovetzky and Yanchuan Chang 2311

PRKV: Page Restructed KV Cache for High Accuracy and Efficiency LLM Generation
Fang Wu, Congming Gao, Weixi Zhu and Jiwu Shu 2355

LIBERO-Para: A Diagnostic Benchmark and Metrics for Paraphrase Robustness in VLA Models
Chanyoung Kim, Minwoo Kim, Minseok Kang, Hyunwoo Kim and Dahuin Jung 2378

Triggering Chain-of-Thought via Latent Feature Interventions in Large Language Models
Zhenghao He, Guangzhi Xiong, Bohan Liu, Sanchit Sinha and Aidong Zhang 2415

MedRoundsQA: A Persona and Difficulty Aware Evaluation for Multi-Turn Medical Consultations
Youssef Mohamed, Ahmed Heakl, Qinrong Cui, Junhong Liang, Rafiq Ali, Bdour Babillie, Nazira Dunbayeva, Lang Gao, Omar Hussein, Ahmed Nada, Ahmed Mohamed Magdy Mohamed, Jinghui Liu, Salman Khan, Imran Razzak, Yuxia Wang and Xiuying Chen 2433

VidOmni-Bench: A Benchmark for Fine-Grained Video Understanding via Spatio-Temporal Event Verification across Complexity and Duration
Changbeen Kim, Junwon Chang, Kipyo Kim, Risa Shinoda, Kuniaki Saito and Donghyun Kim 2459

LAMPEAE: LLM-Augmented Manifold Probing for Adaptive Event Argument Extraction
Qiyi Wei, Wenqian Shang and Bin Zheng 2478

Think, Look, and Revise: Inconsistency-Aware Visual Self-Correction in MLLMs
Yu Cheng, Arushi Goel and Hakan Bilen 2495

DSAgentBench: Can Agents Automate End-to-End Data-Science Workflows in Real Computer Environments?
Mizanur Rahman, Mohammed Saidul Islam, Ridwan Mahbub, Md Tahmid Rahman Laskar, Shafiq Joty and Enamul Hoque Prince 2517

UNISAF-Bench: Intention-Grounded Safety Evaluation for Unified Multimodal Generation Models
Yooju Shin, Chentao Ye, Weitong Ruan and Rahul Gupta 2541

Escaping the Ambiguity Trap: Content-Level Evidence Tracing for Open-Domain Table QA
Yibin Liu, Yewang Chen and Zhe Li 2568

Is my model perplexed for the right reason? Contrasting LLMs' Benchmark Behavior with Token-Level Perplexity
Zoë Prins, Samuele Punzo, Frank Wildenburg, Giovanni Cinà and Sandro Pezzelle 2584

Man Made Language Models? Evaluating LLMs' Perpetuation of Masculine Generics Bias
Enzo Doyen and Amalia Todirascu 2599

LongFlow: Efficient KV Cache Compression for Reasoning Models
Yi Su, Zhenxu Tian, Dan Qiao, Yuechi Zhou, Zheming Yang, Juntao Li and Min Zhang 2629

A^3: Attention-Aware Accurate KV Cache Fusion for Fast Large Language Model Serving
Yuechi Zhou, Yi Su, Jianxin Zhang, Qingrong Xia, Zhefeng Wang, Xinyu Duan, Baoxing Huai, Kehai Chen, Juntao Li and Min Zhang 2646

Temporal Gains, Spatial Costs: Revisiting Video Fine-Tuning in Multimodal Large Language Models
Linghao Zhang, Jungang Li, Yonghua Hei, Sicheng Tao, Song Dai, Yibo Yan, Zihao Dongfang, Weiting Liu, Chenxi Qin, Hanqian Li, Xin Zou, Jiahao Zhang, ShuHang Xun, Haiyun Jiang and Xuming Hu 2662

Harm or Humor: A Multimodal, Multilingual Benchmark for Overt and Covert Harmful Humor
Ahmed Sharshar, Hosam Elgendy, Yasser Rohaim, Saadeldine Eletter and Yuxia Wang 2680

Enabling KV Caching of Shared Prefix for Diffusion Language Models
Younghun Go, Jaehoon Han, Changyong Shin, Chuck Yoo and Gyeongsik Yang 2699

Parallel Tokenizers: Rethinking Vocabulary Design in Cross-Lingual Transfer of Low-Resource Languages
Muhammad Dehan Al Kautsar and Fajri Koto 2718

Retrieval-Augmented Generation with Retrieval-optimized TRA Structure
JungHyoun Kim, Soohyeong Kim, Seok Jun Hwang, Jeonghyeon Park and Yong Suk Choi . 2735

GAC: Noise-Aware Adaptive Mixing for Hybrid SFT-RL Post-Training
Yuelin Hu, Wei Liu, Zhenbo Yu, Zhengxue Cheng and Li Song 2763

Last Layer Logits to Logic: Empowering LLMs with Logic-Consistent Structured Knowledge Reasoning
Songze Li, Zhiqiang Liu, Zhaoyan Gong, Zhongpu Bo, Zhengke Gui, Lei Liang, Huajun Chen and Wen Zhang 2780

Training-Free Character-Length Control in Summarization with Diffusion Language Models
Steffen Castle, Nils Feldhus, Christopher Ebert, Leonhard Hennig and Sebastian Möller 2803

IndicWinoBias: Measuring Occupational Gender Bias and Error Asymmetry in Indian Languages
Naga Sai Ritvik Modumudi, Dipti Sharma and Parameswari Krishnamurthy 2814

Beyond Tokens: Semantic-Aware Speculative Decoding for Efficient Inference by Probing Internal States
Ximing Dong, Shaowei Wang, Dayi Lin, Boyuan Chen and Ahmed E. Hassan 2836

MultiSessionCollab: Learning User Preferences with Memory to Improve Long-Term Collaboration
Shuhaib Mehri, Priyanka Kargupta, Tal August and Dilek Hakkani-Tür 2852

TempCore: Are Video QA Benchmarks Temporally Grounded?
Hyunjong Ok and Jaeho Lee 2880

STARK: Strategic Advantage Reinforcement with Confidence-Driven Exploration for Long-Horizon Agentic Learning
Jinyang Wu, Xinan Chang, Yuhao Shen, Zhengxi Lu, Shuo Yang, Mingkuan Feng, Jiangyan Yi, Shuai Zhang, Zheng Lian, Zhengqi Wen and Jianhua Tao 2902

Attribute-Based Activation Steering of LLMs for Group-Specific Explanation Generation
Leandra Fichtel, Janek Prange and Henning Wachsmuth 2929

How Much Supervised Fine-Tuning Is Enough? Dissecting the Impact of SFT Extent on Downstream Reinforcement Learning Performance
Shirley Kokane 2960

Simulating Multimodal Multiparty Student Interactions with Large Language Models for Training Adaptive Pedagogical Agents
Tristan Maidment, Timothy Nokes-Malach, Diane Litman, Erin Walker and Adriana Kovashka 2972

Mask-Free Privacy Extraction and Rewriting: A Domain-Aware Approach via Prototype Learning
Xiaodong Li, Yuhua Wang, Qingchen Yu, Zixuan Qin, Yifan Sun, Qinnan Zhang, Hainan Zhang and Zhiming Zheng 2987

Helping Figures Tell their Story! Paper-Grounded Video Generation Explaining Complex Scientific Figures
Ishani Mondal, Javad Baghirov and Jordan Boyd-Graber 3017

AuroraCounsel: A Multi-Agent Large Language Model Collaboration Framework for Psychological Counseling
Yuean Zhu, ZiChao Song and Haiming Luo 3052

AfrIFact: Cultural Information Retrieval, Evidence Extraction And Fact Checking for African Languages
Israel Abebe Azime, Jesujoba Oluwadara Alabi, Crystina Zhang, Iffat Maab, Atnafu Lambebo Tonja, Tadesse Destaw Belay, Folasade Peace Alabi, Salomey Osei, Saminu Mohammad Aliyu, Nkechinyere Faith Aguobi, Bontu Fufa Balcha, Blessing Kudzaishe Sibanda, Davis David, Mouhamadane Mboup, Daud Abolade, Neo Putini, Philipp Slusallek, David Ifeoluwa Adelani and Dietrich Klakow 3085

Evo: Autoregressive–Diffusion Large Language Models with Evolving Balance
Junde Wu, Minhao Hu, Jiayuan Zhu, Yuyuan Liu, Tianyi Zhang, Kang Li, Jingkun Chen, Jiazhen Pan, Qirong Ho, Min Xu and Yueming Jin 3108

Long Document Reconstruction Unlocks Scalable Long-Context RLVR
Yao Xiao, Lei Wang, Yue Deng, Guanzheng Chen, Ziqi Jin, Jung-jae Kim, Xiaoli Li, Roy Ka-Wei Lee and Lidong Bing 3122

Learning the Cue or Learning the Word? Analyzing Generalization in Metaphor Detection for Verbs
Sinan Kurtyigit, Sabine Schulte im Walde and Alexander Fraser 3136

Overreliance on AI in Information-seeking from Video Content
Anders Giovanni Møller, Elisa Bassignana, Francesco Pierri and Luca Maria Aiello 3147

HIPO: Instruction Hierarchy via Constrained Reinforcement Learning
Keru Chen, Jun Luo, Sen Lin, Yingbin Liang, Alvaro Velasquez, Nathaniel D. Bastian and Shaofeng Zou 3168

Evaluating Communicative Belief Updates in Large Language Models via Implicature Recognition and Cancellation
Cesare Spinoso-Di Piano, Verna Dankers, Marius Mosbach and Jackie Chi Kit Cheung 3192

DiscoTrace: Representing and Comparing Answering Strategies of Humans and LLMs in Information-Seeking Question Answering
Neha Srikanth, Jordan Boyd-Graber and Rachel Rudinger 3218

Micro Language Models Enable Instant Responses
Wen Cheng, Tuochao Chen, Karim Helwani, Sriram Srinivasan, Luke Zettlemoyer and Shyamnath Gollakota 3240

Scaling Laws for Mixed Quantization
Zeyu Cao, Yao Lai, Boyang Gu, Cheng Zhang, Pedro Gimenes, Jianqiao Lu, Xitong Gao, Jianyi Cheng, Robert D. Mullins and Yiren Zhao 3265

The Curious Case of Factual (Mis)Alignment between LLMs' Short- and Long-Form Answers
Saad Obaid Ul Islam, Anne Lauscher and Goran Glavaš 3309

Audio-Reward-R1: Reasoning Reward Model for Unified Audio Understanding and Generation
Hao Gu, Xiaocheng Yang, Jinyang Wu, Hongbo Zhao, Yunzhe Hao, Songcheng Xu, Ye Bai, Zhao You and Shidong Shang 3327

VTCBench: Can Vision-Language Models Understand Long Contexts with Vision-Text Compression?
Hongbo Zhao, Meng Wang, Fei Zhu, Wenzhuo Liu, Bolin Ni, Fanhu Zeng, Gaofeng Meng, Zhaoxiang Zhang and Cheng-Lin Liu 3349

When Hate Meets Facts: LLMs-in-the-Loop for Check-worthiness Detection in Hate Speech
Nicolás Benjamín Ocampo, Tommaso Caselli and Davide Ceolin 3380

Keeping Eyes on the Image: Dynamic Visual Attention Recalibration for Vision-Language Reasoning
Yuna Jeong, Junhyeong Park, Mingyu Kang, Hyein Seo, Keon Kim and Yong Suk Choi 3398

MAPLE: Masked Adapter Prototype Learning for OOD generalization
Adarsh Kappiyath, Abhra Chaudhuri, Jiaxi Li, Ajay Kumar Jaiswal, Xilu Wang, Li Shen, Souvik Kundu, Xiatian Zhu and Lu Yin 3419

From Test-taking to Cognitive Scaffolding: A Pedagogical Diagnostic Benchmark for LLMs on English Standardized Tests
Luoxi Tang, Tharunya Sundar, Yuqiao Meng, Shuai Yang, Ankita Patra, Lakshmi Manohar Chippada, Jiqian Zhao, Yi Li, Weicheng Ma and Zhaohan Xi 3433

JuDi: Revisiting Judge Decoding from First Principles via Training-Free Distributional Divergence
Shengyin Sun, Yiming Li, Renxi Liu, Weizhe Lin, Hui-Ling Zhen, Xianzhi Yu, Mingxuan Yuan and Chen Ma 3471

Demystifying Hybrid Thinking: Can LLMs Truly Switch Between Think and No-Think?
Shouren Wang, Wang Yang, Xianxuan Long, Debargha Ganguly, Qifan Wang, Vipin Chaudhary and Xiaotian Han 3491

TKG-Thinker: Towards Dynamic Reasoning over Temporal Knowledge Graphs via Agentic Reinforcement Learning
Zihao Jiang, Miao Peng, Zhenyan Shan, Wenjie Xu, Ben Liu, Gong Chen, Ziqi Gao and Min Peng 3512

Beyond Junk Tokens: Subsumed Tokens in Japanese Subword Tokenization
Toshiki Yamaguchi, Kazuo Hara and Ikumi Suzuki 3534

SAFEGuard: Detect Optimization-Based Jailbreak Atacks Through Harmful Semantic Analysis and Fluency Measurement
Quoc Viet Vo, Trung Le, Damith Ranasinghe and Ehsan Abbasnejad 3546

HIPPO: Accelerating Video Large Language Model Inference via Holistic-aware Parallel Speculative Decoding
Qitan Lv, Tianyu Liu, Wen Wu, Xuenan Xu, Bowen Zhou, Feng Wu and Chao Zhang 3574

MIMO: Multilingual Information Retrieval via Monolingual Objectives
Youngjoon Jang, Seongtae Hong and Heuiseok Lim 3597

Verification Trap: Understanding Test-Time Selection Failures under False Premises in Code Generation
Feng He, Hejia Wang, Linghao Meng, Ming Gao and Qiankun Li 3613

W2T: LoRA Weights Already Know What They Can Do
Xiaolong Han, Ferrante Neri, Jiang Zijian, Fang Wu, Yanfang Ye, Lu Yin and Zehong Wang3638

On the Optimality of Kinship Naming: an Information-theoretic Approach
Phong Le, Mees Lindeman and Raquel G. Alhama 3654

From Chatbots to Confidants: A Cross-Cultural Study of LLM Adoption for Emotional Support
Natalia Amat-Lefort, Mert Yazan, Amanda Cercas Curry and Flor Miriam Plaza-del-Arco . 3668

To Answer or to Abstain: Mitigating Search-Agent Hallucinations via Abstention-Aware Reinforcement Learning
Fengji Zhang, Tianyu Fan, Yuxiang Zheng, Xinyao Niu, Chengen Huang, Jacky Keung and Bei Chen 3696

Predicting Multilingual Classification and Translation Performance of LLMs with Cross-Lingual Alignment – Is English Enough?
Adnan Al Ali, Kathy Hämmerl, Jindřich Libovický and Alexander Fraser 3719

Preference-Aligned Data Synthesis for Fine-Grained Intent Detection
Sitong Shen and Xiaotong Zhang 3742

M-QUEST - Meme Question-Understanding Evaluation on Semantics and Toxicity
Stefano De Giorgis, Ting-Chih Chen and Filip Ilievski 3759

More Test-Time Compute Can Hurt: Overestimation Bias in LLM Beam Search
Gal Dalal, Assaf Hallak, Gal Chechik and Yftah Ziser 3785

AgentFold: Closed-Loop Agentic Search for Protein Folding Model Design
Mingquan Liu, Jiangyu Chen, Hanqun Cao, Xujun Zhang, Pengsen Ma, Xiangru Tang, Shuting Jin, Annie Zheng, Zhuo Yang, Tianfan Fu, Fang Wu and Xiangxiang Zeng 3802

LDC: Learning to Generate Research Idea with Dynamic Control
Ruochen Li, Liqiang Jing, Chi Han, Jiawei Zhou and Xinya Du 3831

Evaluating Style-Personalized Text Generation: Challenges and Directions
Anubhav Jangra, Bahareh Sarrafzadeh, Silviu Cucerzan, Adrian de Wynter and Sujay Kumar Jauhar 3854

DSPA: Dynamic SAE Steering for Data-Efficient Preference Alignment
James Wedgwood, Aashiq Muhamed, Mona T. Diab and Virginia Smith 3879

Generalizable Lifelong Model Editing via Preference Optimization
Dahyun Jung, Suhyune Son and Heuiseok Lim 3899

Fight Poison with Poison: Enhancing Robustness in Few-shot Machine-Generated Text Detection with Adversarial Training
Wenjing Duan, Qi Zhou and Yuanfan Li 3921

When Denser Credit Is Not Enough: Evidence-Calibrated Policy Optimization for Long-Horizon LLM Agent Training
Yuanfan Li, Qi Zhou, Wenjing Duan and Lu Chen 3948

OffSeeker: Online Reinforcement Learning Is Not All You Need for Deep Research Agents
Yuhang Zhou, Kai Zheng, Qiguang Chen, Mengkang Hu, Qingfeng Sun, Can Xu and Jingjing Chen . . . 3966

Dental-TriageBench: Benchmarking Multimodal Reasoning for Hierarchical Dental Triage
Ziyi He, Yushi Feng, Shuangyu Yang, Yinghao Zhu, Xichen Zhang, Tai Pak Chuen Patrick, Lo Hei Yuet, Songying Wu, Weifa Yang and Lequan Yu . . . 3987

Hyperspherical Semantic Expansion via von Mises–Fisher Mixtures for Zero-Shot Dense Retrieval
Sehong Oh, DongHui Kim and Junseok Kwon . . . 4013

SAGE: Semantic Anchor-Guided Evolution for Grounded Medical QA Data Synthesis
Chuan Li, Chengyu Wang, Cen Chen, Ye Lyu, Mingyuan Fan and Ming Gao . . . 4027

ECHO: Early-layer Collaborative Hierarchical Orchestration with Bonus Logits in Speculative Decoding
Ziyang Ma, Zihong Zhang, Zuchao Li, Lefei Zhang, Baoyuan Qi, Siqi Li and Simin Yu . . . 4047

On the Context Sensitivity of LLM Moral Judgment
Adrian Sauter and Mona Schirmer . . . 4068

GeoMotionGPT: Geometry-Aligned Motion Understanding with Large Language Models
Zhankai Ye, Bofan Li, Yukai Jin, Shuoqiu Li, Wei Wang, Yanfu Zhang, Shangqian Gao and Xin Liu . . . 4114

SustainableQA: A Comprehensive Question Answering Dataset for Corporate Sustainability and EU Taxonomy Reporting
Mohammed Ali, Abdelrahman Abdallah and Adam Jatowt . . . 4130

HyLaT: Efficient Multi-Agent Communication via Hybrid Latent–Text Protocol
Xinyi Mou, Siyuan Wang, Zejun Li, Yulan He and Zhongyu Wei . . . 4153

DebugLM: Learning Traceable Training Data Provenance for LLMs
Wenjie Jacky Mo, Qin Liu, Xiaofei Wen, Wenxuan Zhou, Zhe Zhao and Muhao Chen . . . 4168

DiscoSign: Discourse-Aware Text to Sign Language Gloss Translation
Vasileios Baltatzis, Mert Inan, Connor Gillis, Raja Kushalnagar, Lorna Quandt, Leah Findlater and Colin Lea . . . 4183

CRISP: Cliff-awaRe Input-adaptive Sparse Prefilling with Structural-Mass-Motivated Routing
Huy Huu Nguyen, Chien Van Nguyen, Franck Dernoncourt, Ryan A. Rossi, Linh Ngo Van, Jieyang Chen and Thien Huu Nguyen . . . 4199

Response-Aware User Memory Selection for LLM Personalization
Jillian Fisher, Jennifer Neville and Chan Young Park . . . 4215

Do Latent Tokens Think? A Causal and Adversarial Analysis of Chain-of-Continuous-Thought
Yuyi Zhang, Boyu Tang, Tianjie Ju, Sufeng Duan and Gongshen Liu . . . 4268

MAS-on-the-Fly: In-Context Structural Adaptation of LLM-Based Multi-Agent Systems
Guangyi Liu, Linhaojun, Zenghuan, Heng Wang and Quanming Yao . . . 4282

Learning an Efficient Multi-Turn Dialogue Evaluator from Multiple LLM Judges
Yuqi Tang, Kehua Feng, Yunfeng Wang, Zhiwen Chen, Chengfei lv, Gang Yu, Qiang Zhang, Keyan Ding and Huajun Chen . . . 4304

Safety-Aware Contrastive Decoding for Contextual Safety in Multimodal Large Language Models
Zheyuan Liu, Zhangchen Xu, Guangyao Dou, Xiangchi Yuan, Zhaoxuan Tan, Radha Poovendran and Meng Jiang 4328

ReBOL: Retrieval via Bayesian Optimization with Batched LLM Relevance Observations and Query Reformulation
Anton Korikov and Scott Sanner 4364

FDARxBench: Benchmarking regulatory and clinical reasoning on FDA generic drug assessment
Betty Xiong, Jillian Fisher, Benjamin Newman, Meng Hu, Shivangi Gupta, Yejin Choi, Lanyan Fang and Russ B Altman 4377

SIRI: Self-Internalizing Reinforcement Learning with Intrinsic Skills for LLM Agent Training
Zhongyu He, Yuanfan Li, Fei Huang, Tianyu Chen, Siyuan Chen, Xingyang Li, Meng Hsuan Yu, Xiangrong Liu, Leyi Wei, Lu Pan, Ke Zeng and Xunliang Cai 4402

When i.i.d. Factorization Breaks Language: The Structural Failure of Discrete Diffusion Language Models
Hyukhun Koh, Seongho Joo, Haesung Pyun and Kyomin Jung 4417

From Atomic to Agentic: Towards Interpretable Evaluation of LLMs' Agentic Mathematical Capabilities
Jiayi Kuang, Yinghui Li, Yunze Song, Keyu Chen, Zhifeng Shen, Yangning Li, Yidong Wang, di Yin, Ruizhi Qiao, Xing Sun, Kai Jin, Ying Shen, Liang Lin and Philip S. Yu 4437

Decoupled Contrastive Decoding via Expert-Aligned Drafting
Zhixuan Liu, Zhichen Dong, Yuanfu Wang and Chao Yang 4466

ChartAttack: Testing the Vulnerability of LLMs to Malicious Prompting in Chart Generation
Jesus-German Ortiz-Barajas, Jonathan Tonglet, Vivek Gupta and Iryna Gurevych 4495

CyclicJudge: Mitigating Judge Bias Efficiently in LLM-based Evaluation
Ziyi Zhu, Olivier Tieleman, Alexey Bukhtiyarov and Jinghong Chen 4527

Beyond Top-k : Mass-Aware Support Sampling for On-Policy Distillation
Yu Wang, Xinran Chen, Yuchen Li, Zhang Honghai, Rui Kong, Jiamin Chen, Lingyong Yan, Linghe Kong, Chen Ma, Haoyi Xiong and Dawei Yin 4540

Stabilizing Unsupervised Self-Evolution of MLLMs via Continuous Softened Retracing reSampling
Yunyao Yu, ZhengXian Wu, Zhuohong Chen, Hangrui Xu, Zirui Liao, Xiangwen Deng, Zhifang Liu, Senyuan Shi and Haoqian Wang 4555

Mitigating Model Merging Conflicts via Knowledge-Aware Task Vector Purification
Bang An, Yibo Yang, Philip Torr and Bernard Ghanem 4572

Medical Bilingualism: Cross-Ontology Graph Retrieval for Incommensurable Knowledge Bases
Zixuan Wang, Ke Lu, Zihan Wang, Junwei Lei, Junhao Dong, Jinyu Guo and Chaomeng Chen 4591

TTS-Hub: Leveraging Modular LoRAs and Arithmetic Composition for Controllable Text-to-Speech
Xiang Li, Shiqi Zhang, Zichen Xu, Wenyuan Gu, Hongru Xiao, Bo Cheng, Li Zhou, Jiale Han and Benyou Wang 4613

Share More, Search Less: Collaborative Parallel Thinking for Efficient Test-Time Scaling
Xinglin Wang, Hao Lin, Shaoxiong Feng, Peiwen Yuan, Yiwei Li, Jiayi Shi, Yueqi Zhang, Chuyi Tan, Ji Zhang, Boyuan Pan, Yao Hu and Kan Li 4640

Why Is This Joke Offensive? Interpreting Offensiveness in Chinese Dark Humor Through Knowledge and Rhetoric
Jiecheng Jiang, Herun Wan, Zihang Fu, Hai-Tao Yu, Min-Yen Kan and Jiaying Wu 4675

The Facade of Truth: Uncovering and Mitigating LLM Susceptibility to Deceptive Evidence
Herun Wan, Jiaying Wu, Minnan Luo, Fanxiao Li, Zhi Zeng and Min-Yen Kan 4697

Inferring the Unspoken: Aligning Embodied Agents with Implicit Preferences
Manjie Xu, Xinyi Yang, Wei Liang, Chi Zhang and Yixin Zhu 4725

Anticipating User Interpretations in Emotional Support Conversations via Inference-Time User-Side Simulation
Jiaqi Liu, Jiakang Xu, Zhongqiang Du, Haonan Zheng, Yankun Yang and Wenbin Jiang 4746

H-OPD: Confidence Aware Heterogeneous Multi-Teacher Multimodal On-Policy Distillation
Qixiang Yin, Huanjin Yao, Cai Yuchen, Jianghao Chen, Ziyi Wang, Min Yang, Fei Su and Zhicheng Zhao 4780

Causal Probing for Internal Visual Representations in Multimodal Large Language Models
Zehao Deng, Tianjie Ju, Zheng Wu, Liangbo He, Jun Lan, Huijia Zhu, Weiqiang Wang and Zhuosheng Zhang 4798

GradSentry: Gradient Spectral Entropy for Backdoor Sample Filtering in Large Language Model Fine-Tuning
Haodong Zhao, Tianyi Xu, Tianhang Zhao, Zhuosheng Zhang and Gongshen Liu 4820

SABER: Benchmarking Operational Safety of LLM Coding Agents in Stateful Project Workspaces
Qi HU, Yifeng Tang, Qinghua Wang, Lanyang Zhao, Pengji Zhang, Yuhao Qing, Xin Yao, Dong Huang, Lin Zhang and Zhuoran Ji 4842

HyperProve: Answer-Guided Hypergraph Expansion for Multi-Hop Question Answering
Nguyen Phu An, Dung Nguyen Quang, Luu Hieu An, Linh Ngo Van, Trung Le and Thien Huu Nguyen 4867

From Sufficiency to Reflection: Reinforcement-Guided Thinking Quality in Retrieval-Augmented Reasoning for LLMs
Jie He, Victor Gutierrez Basulto and Jeff Z. Pan 4883

AnInter: Anti-Interference Training for Long Context Retrieval
Lijie Ye 4909

LLMs Regret Before They Say It: Early Detection and Compositional Architecture of Regret in Hidden States
Xiangxiang Cui, Shu Yang, Tianjin Huang, Wanyu Lin, Lijie Hu and Di Wang 4923

Structural Anchor Pruning: Training-Free Multi-Vector Compression for Visual Document Retrieval
Zhuchenyang Liu, Ziyu Hu, Yao Zhang and Yu Xiao 4947

Causal Tongue-Tie: LLMs Can Encode Causal Direction, But Their Yes/No Outputs Fail to Express
Ziyi Ding and Xiao-Ping Zhang 4976

You Only Align Once: Propagating Cooperative Behaviors in Multi-Agent Systems through Seed Agents
Nicole Summer Hsing, Asuka Yuxi Zheng, Yi Zhao, Haoqin Tu and Jen-tse Huang 4999

CoDiQ: Test-Time Scaling for Controllable Difficult Question Generation
Zhongyuan Peng, Caijun Xu, Changyi Xiao, Shibo Hong, Ge Zhang, Wenhao Huang and Yixin Cao 5037

HaluMem: Evaluating Hallucinations in Memory Systems of Agents
Ding Chen, Simin Niu, Li Kehang, Peng Liu, Xiangping Zheng, Bo Tang, Xinchi Li, Feiyu Xiong and Zhiyu li 5073

FormuEvo: LLM-Guided Evolution for Discovering Solver-Efficient Mixed-Integer Programming Formulations
Haofeng Yuan, Jianing Peng, Jieyi Bi, Ni Zhang, Shiji Song and Zhiguang Cao 5109

UCO: A Multi-Turn Interactive Reinforcement Learning Method for Adaptive Teaching with Large Language Models
Shouang Wei, Min Zhang, Xin Lin, Bo Jiang, Kun Kuang, Jingyuan Chen and Zhongxiang Dai 5136

ST-EVO: Towards Generative Spatio-Temporal Evolution of Multi-Agent Communication Topologies
Xingjian Wu, Xvyuan Liu, Junkai Lu, Siyuan Wang, Xiangfei Qiu, Yang Shu, Jilin Hu, Chenjuan Guo and Bin Yang 5156

Thinking Without Speaking: Efficient Tree of Thoughts Reasoning via Latent Steering
Zhanshuo Dong, Zhenyu Wen, Xiong Li, Jie Su, Jianbin Qin, Ye Yuan and Zhen Hong 5182

Investigating and Enhancing Concept Induction in LLMs: A Prototype-based Property-augmented Approach
Mingyang Lyu, Bin Zhang, Chenlong Bao, Jin-Tao Tang, Shasha Li and Ting Wang 5203

How Robust is OCR-Reasoning? Evaluating OCR-Reasoning Robustness of Vision-Language Models under Visual Perturbations
Yuxing Cheng, Yuan Wu and Yi Chang 5220

LongSpace: Exploring Long-Horizon Spatial Memory from Perception to Recall in Video
Shiqiang Lang, Jing Liu, Haoyang He, Peiwen Sun, Yuanteng Chen, Tao Liu, Lan Yang, Longteng Guo and Honggang Zhang 5239

SUGAR: Spectral and Geometry-Aware Alignment for Matryoshka Representation Distillation
Hai An Vu, Phung Gia Huy, Minh-Phuc Truong, Linh Ngo Van, Thanh Hong Nguyen, Thanh-Toan Do and Trung Le 5257

EarlySelectKV: Anticipatory KV Selection via Cross-Layer Query Proxies
Yuhao Liu, Yuehao Li, Linghe Zhang, Gaobin Huang, Junran Li, Hongchang Wang and Haifang Jian 5272

IntentVLA: Short-Horizon Intent Modeling for Aliased Robot Manipulation
Shijie Lian, Bin Yu, Xiaopeng Lin, Zhaolong Shen, Laurence Tianruo Yang, Yurun Jin, Haishan Liu, Changti Wu, Hang Yuan, Cong Huang and Kai Chen 5292

VASTU: Language Models Struggle to Recognize Online Community Values
Agam Goyal, Xianyang Zhan, Charlotte Lambert, Koustuv Saha and Eshwar Chandrasekharan 5311

RW-TTT: Batched Serving for Request-Owned Test-Time Training State
Jian Yang, Zhizhuo Kou, Chen Han, Hao Zhang, Yao Tian, Sirui Han and Yike Guo 5328

Beyond State Consistency: Behavior Consistency in Text-Based World Models
Youling Huang, Guanqiao Chen, Junchi Yao, Lu Wang, Fangkai Yang, Chao Du, ChenZhuo Zhao, Pu Zhao, Qingwei Lin, Saravan Rajmohan and Dongmei Zhang 5339

Identity-Preserving Detoxification via Local Causal Effect Ranking on Masked Diffusion Language Models
Yucheng Lin, Yue Liu and Jiaojiao Jiang 5360

ParaBridge: Bridging Paralinguistic Perception and Dialogue Behavior in Speech Language Models
Yuxiang Wang, Qinke Ni, Shengbo Cai, Wan Lin, Liqiang Zhang and Zhizheng Wu 5379

You Can't Trust Your Agent: A Survey on Backdoor Threats and Defenses in Intelligent Agents
Long Tuan Ho, Yahui Ding, Huimin Xu, Shuai Zhao and Anh Tuan Luu 5403

Train Smarter, Not Harder: Switching Signal-Guided Training in Active Learning
Nagham Omar, Maya Rozenshtein, Evgeny Mishlyakov and Avigdor Gal 5423

Learning to Reason Before Masking for Referring Audio-Visual Segmentation
Zhixuan Wu, Ye Ni, Yiming Chen, Wenyuan Gu and Bo Cheng 5446

LookStep: Efficient Vision-Language Navigation with Linguistic Foresight and Event Driven Memory
Kun-Yang Yu, Yingzhe Li, Hongyu Xu, Shi-Yu Tian, Zhi Zhou, Yang Chen, Ming Yang, Sheng Wang, Qing Yu, Lan-Zhe Guo and Yu-Feng Li 5465

Q-Align: Query-Boundary Multilingual Alignment for LLM Safety Guardrails
Jiaqi Weng, Meizhen Liu, Xu Guo, Junyu Lin, Jinfeng Li, Jialing Tao, YueFeng Chen and Hui Xue 5484

How Should LLMs Consume High-Quality Data? Optimal Data Scheduling via Quality-Aware Functional Scaling Laws
Zhitao Zhu, Xili Wang, Shizhe Wu, Jiawei Fu and Xiaoqing Liu 5499

DP-ES: Differentially Private Evolution Strategies for Prompt Optimization
Ziniu Liu, Aiping Li, Yue Han, Han Yu, Junjian Zhang, Dong Zhu, Changjian Li and Shiqiang Zhang 5529

Teaching LLMs to Abstain via Fine-Grained Semantic Confidence Reward
Hao An and Yang Xu 5546

Streaming Communication in Multi-Agent Reasoning
Zhen Yang, Xiaogang Xu, Wen Wang, Cong Chen, Xander Xu and Ying-Cong Chen 5567

When Surprise Is Not Enough: Validation-Aware Rewards for LLM-Guided Scientific Discovery
Yifan Gao, Haoyue Li, Feng Yuan and Xin Gao 5596

MCA: Modality Composition Awareness for Robust Composed Multimodal Retrieval
Qiyu Wu, Shuyang Cui, Satoshi Hayakawa, Wei-Yao Wang, Hiromi Wakaki and Yuki Mitsufuji 5612

ClinHallu: A Benchmark for Diagnosing Stage-Wise Hallucinations in Medical MLLM Reasoning
Sicheng Yang, Hangjie Yuan, Wenjun Zhang, Jinwang Wang, Yichen Qian, Weihua Chen, Fan Wang and Lei Zhu 5629

Ready to Speak: Aligning LLMs for TTS-Friendly Text Generation
Thibaut Thonet, Jos Rozen and Laurent Besacier 5650

Efficient Temporal-aware Matryoshka Adaptation for Temporal Information Retrieval
Tuan-Luc Huynh, Weiqing Wang, Trung Le, Thuy-Trang Vu, Dragan Gasevic, Yuan-Fang Li and Thanh-Toan Do 5677

IndicParam: Benchmark to evaluate LLMs on low-resource Indic Languages
Ayush Maheshwari, Kaushal Sharma, Vivekkumar Vasudevbhai Patel and Aditya Maheshwari 5702

PVFH-Bench: Benchmarking Premise-Induced Visual Fixation Hallucinations in Vision-Language Models
Xinhao Wang, Junbo Cui, Zhenghao Liu, Yuan Yao, Xiaohua Li, Yu Gu and Ge Yu 5722

Semantic Self-Guided Watermarking with Enhanced Text Quality for Large Language Models
Linkang Yang, Bo Zhang, Bin Mao, Jiajun Cheng, Junbo Fu, Bingxu Han, Bo Han, Jun Zhang, Guoshuai Zhao, Zhunchen Luo, Yunbo Cao, Xiaoying Bai and Guotong Geng 5740

Long-to-Short Video Evidence Reasoning for Grounded Question Answering
Kaiyan Chen, Junbin Xiao and Xun Yang 5752

Beyond Single-Policy: Evaluating Composed Organization-Specific Policy Alignment in LLM Chatbots
Yingjie Liu, Yongxiang Hu, Xuan Wang, Yilun Li, Yunlei Wei, Xiaoyu Wang and Yangfan Zhou 5767

From Accuracy to Robustness: A Study of Rule- and Model-based Verifiers in Mathematical Reasoning
Yuzhen Huang, Weihao Zeng, Xingshan Zeng, Qi Zhu and Junxian He 5787

From Gaps to Active Retrieval : Open-Domain Iterative Timeline Completion via Historical Trajectory Analogy and Dynamic Ontology
Yunhui Zheng, Chenlong Bao, Shasha Li, Ting Wang and Jin-Tao Tang 5809

PGA-KD: Principal Geometry Alignment and Semantic Consistency for VLM2Vec Distillation
The-Anh Tran, Thanh Xuan Nguyen, Duc Anh Nguyen, Dinh Viet Sang, Linh Ngo Van and Thien Huu Nguyen 5827

Fully Unleashing the Multimodal Attacker: Meta-Adaptive Jailbreaking of Vision-Language Models
Shen Pang, Benlei Cui, Yuke Wang, Xuemei Dong, Yuwen Zhai, Jingqun Tang, Haiyang Yu, Hui Xue, Longtao Huang and Haiwen Hong 5848

GGSS: Geodesic-Gated Spherical Steering for Inference-Time Debiasing of Generative Vision-Language Models
Yiqun Sun, Junyu Chen, Pengfei Wei and Lawrence B. Hsieh 5875

Validity Is Not Visibility: A Lifecycle FSM for Agent Memory
Shuo Zhou, Pengcheng Xu, Shaotong qi, Li Que, Ming Sun and Han Li 5895

DocFollow: A Hierarchical Fine-Grained Constraints Following Benchmark in Long Contexts for LLMs
Yanji He, Zhao Yang, Yuxin Jiang, Yiqian Yang, Jiaheng Wei and Wei Wang 5914

Moral Lenses, Political Coordinates: Towards Ideological Positioning of Morally Conditioned LLMs
Chenchen Yuan, Bolei Ma, Zheyu Zhang, Bardh Prenkaj, Frauke Kreuter and Gjergji Kasneci 5950

MangaFlow: An End-to-End Agentic Framework for Controllable Story to Manga Generation
Muyao Wang, Zeke Xie, Yanhao Chen, Lixin Xiu and Hideki Nakayama 5985

LSC-Pattern: Interpretable Lexical Semantic Change Analysis via Diachronic Sense Distributions
Zhidong Ling and Mamoru Komachi 6003

LLM Feedback for Topic Model with Preference Optimization and Contrastive Learning
Tung Nguyen, Dat Phan Trong, Linh Ngo Van, Duc Anh Nguyen, Dinh Thai Hoang and Diep Nguyen 6021

Are Android GUI Agents Robust Against Runtime Anomalies? AnTrap: Evaluating Agents in Dynamic Adversarial Environments
Guo Gan, Yilun Zhao, Cong Chen, Jinbiao Wei, Tingyu Song, Zheyuan Yang, Lin Fu and Hong Zhou 6046

EmbodiChat: Towards Zero-Latency Text-driven Talking Face Generation for Embodied Dialogue
Tiecheng Zhang, Huan Yu, Zhu Runjie, Zequan Xu, Wei Wang, Kaixiang Huang and Jin Wang 6070

Agent-G^2: Gaussian Guidance for Agentic Reinforcement Learning
Zixuan Wang, Yanrui Miao, Zhengxi Lu, Teng Pan, Yiwen Qiu, Hongxing Li, Peng Qiu, Ruiqing Zhang and Yongliang Shen 6096

AdaCode: Adaptive Lossless Compression for BFloat16 KV Cache
Changshi Li, Peiyu Liu, Liang Chen, Yuyao Zhai and Jerry Zhijian Yang 6112

MemDreamer: Decoupling Perception and Reasoning for Long Video Understanding via Agentic Hierarchical Graph Memory
Cong Chen, Guo Gan, Kaixiang Ji, Zhaoyang Zhang, Zhen Yang, Guangming Yao, Hao Chen, Jingdong Chen, Yi Yuan and Chunhua Shen 6131

PrimeTTT: Priming Test-Time Training for LLMs via Meta-Learned Initialization
Xiangqi Wang, Dilinuer Aishan, Yujun Zhou, Yue Huang and Xiangliang Zhang 6152

CapRiCorn-1K: A Comprehensive Benchmark for Video Captioning and Subject Referential Consistency Across Temporal Scales
Xinlong Chen, Jiafu Tang, Yue Ding, Yizhuo Jia, Bozhou Li, Bohan Zeng, Yang Shi, Shihao Li, Yiyan Ji, Qiang Liu, Weihong Lin, Yuanxing Zhang, Pengfei Wan, Liang Wang and Tieniu Tan . . 6171

Soft Tokens Are Off-Manifold: Geometric Correction for Continuous-Space Reasoning in Large Language Models
Jing Gao, Weiqiang Jin, Yunqi Mi, Peigen Ye, Hongyang Du, Biao Zhao, Yimin Deng, Chengxu Liu, Xueming Qian and Guoshuai Zhao 6198

XRepoTest: Benchmarking Multilingual Repository-Level Unit Test Generation for Large Language Models
Dung Le Quang, Dong Cao Van, Nam Le Hai, Linh Ngo Van, Thi-Mai-Anh Bui and Phuong T. Nguyen 6224

OmniV-Med: A Scalable Training Recipe for Unified Medical Vision-Language Understanding
Songtao Jiang, Yuan Wang, Chenyi Zhou, Tianxiang Hu, Fucheng Niu and Zuozhu Liu 6263

Auditing Evidence Use in Medical LLM Diagnosis
Junchi Liao, Jiawen Deng and Fuji Ren 6280

ForesightSafety-SAGE: A Fully Automated Scenario Generation and Safety Evaluation Framework for LLM Agents
Lu Jia, Haibo Tong, Feifei Zhao, Jindong Li, Dongqi Liang, Ping Wu, Qian Zhang and Yi Zeng 6302

Train Overcomplete, Deploy Compact: Scaling Recovery Capacity for Structured LLM Pruning
Seungmin Oh, Donggeon Lee and Jongbin Ryu 6336

NOSA: Native and Offloadable Sparse Attention
Yuxiang Huang, Pengjie Wang, Jicheng Han, Weilin Zhao, Zhou su, Ao Sun, Hongya Lyu, Hengyu Zhao, Yudong Wang, Chaojun Xiao, Xu Han and Zhiyuan Liu 6356

HarmReduction: Benchmarking LLMs in Harm Reduction Information Provision to Support People Who Use Drugs
Kaixuan Wang, Chenxin Diao, Jason T. Jacques, Zhongliang Guo, Loraine Clarke and Shuai Zhao 6382

SkillsVote: Lifecycle Governance of Agent Skills from Collection, Recommendation to Evolution
Hongyi Liu, Haoyan Yang, Tao Jiang, Bo Tang, Feiyu Xiong, Yuyu Luo and Zhiyu li 6401

Beyond Prefix Answering: Parallel Infilling Reading for Long Contexts with Diffusion Language Models
Han Peng, Zican Dong, Peiyu Liu, Jingyuan Zou, Pengfei Li and Xin Zhao 6463

Syntactic Prediction through Reinforcement Learning
Jinman Zhao, Yining Wang, JIahe Liu, Xueyan Zhang, Linbo Cao, Jiakang Huang, Yiren Zhao, Renyi Cai, YiTian Ding and Gerald Penn 6477

Beyond Visual Memory: Mechanistic Diagnostics of Latent Visual Reasoning
Jiawei Guo, Yu Chen, Xiang Wang, Shuai Li, Xinpei Zhao, Huaxing Liu, Shuai Dong, Feifei Zhai and Yu Zhou 6493

HardGen: Generating Hard Samples for Tool-use Agents
Bingguang Hao, Zengzhuang Xu, Yuntao Wen, Xinyi Xu, Quanen Sun, Yang Liu, Zihao Huang, Tong Zhao, Long Chen, Dong Wang, Yicheng Chen, Maolin Wang, Cunyin Peng, Xiangyu Zhao, Chenyi Zhuang and Ji Zhang 6520

FPCO-Dialog: A Multi-Turn False-Premise Benchmark for Correction and Cooperation in Vision-Language Models
Jiayuan Ma, Yuqi Lu, Weiyang Guo, Chenrui Wang, Junyi Shu, Xuebo Liu, Min Zhang and Jing Li 6547

Beyond Task Boundaries: Instruction-Level Parameter Isolation for Supervised Fine-Tuning
Zekai Lin, Long Xiang, Di Liang, Peiyang Liu, Zhanghao Hu, Mengze Hong, Xianjie Wu and Minlong Peng 6571

Frequency Matters: Fast Model-Agnostic Data Curation for Pruning and Quantization
Francesco Pio Monaco, Elia Cunegatti, Flavio Vella and Giovanni Iacca 6596

BeyondAtoms: Benchmarking and Verifying Relational Factuality Beyond Atomic Facts
Zeyang Liu, Jingfeng Xue, Jiarun Fu, Wenbiao Du, Xiuqi Yang, Yifeng Fu, Junbao Chen and Yong Wang 6620

Grammar of the Wave: Towards Explainable Multivariate Time Series Event Detection via Neuro-Symbolic VLM Agents
Sky Chenwei Wan, Yifei Wang, Tianjun Hou, Xiqing Chang and Aymeric Jan 6638

TEMPO: Realistic Multi-Domain Benchmark for Temporal Reasoning-Intensive Retrieval
Abdelrahman Abdallah, Mohammed Ali, Muhammad Abdul-Mageed and Adam Jatowt 6657

Towards Efficient Reasoning: Learning Causal Shortcuts for Diffusion Language Models
Dian Jin, Kairong Han, Baohong Li, Xinpeng Dong, Zijing Hu, Nuanqiao Shan, Fei Wu and Kun Kuang 6710

Right Answers for the Right Reasons: Coherence-Enhanced Policy Optimization for Large Language Models
Xiangqi Wang, Yue Huang, Yujun Zhou, Xiaonan Luo, Kehan Guo, Yuchen Ma and Xiangliang Zhang 6729

Beyond Retrieval: Structuring Evolving and Inconsistent External Knowledge with Proposition Relations for RAG
Noriki Nishida, Fei Cheng and Yuji Matsumoto 6749

A Survey on Representing Linguistic Style: Challenges and Opportunities
Anna Wegmann, Cristina Aggazzotti, Rafael Alberto Rivera Soto and Dong Nguyen 6772

KoALa-Bench: Evaluating Large Audio Language Models on Korean Speech Understanding and Faithfulness
Jinyoung Kim, Hyeongsoo Lim, Eun Seo Seo, Min Ho Jang, Keunwoo Choi, Seungyoun Shin and Ji Won Yoon 6807

FBV-KGQA: Forward-Backward Verification for Knowledge Graph Question Answering
Jiaqi Wu, Jiakang Li, Zhenxuan Ma and Li Wang 6845

YARD: Y-Architecture Register Decoding for Efficient Hallucination Mitigation in Large Vision-Language Models
Ting Chen, Geng Li, Guohao Chen, Yu Hu, Guan Huang, Mai Chen, Langsheng Lei and Yongyong Lin 6864

ContextPilot: Teaching Agents for Proactive Context Management via Fine-grained RL
Zhuoshi Pan, Qizhi Pei, Junru Lu, Honglin Lin, H. Vicky Zhao, di Yin and Xing Sun 6886

Need to Know: Contextual-Integrity-Grounded Query Rewriting for Privacy-Conscious LLM Delegation
Xinyue Huang, Wenyuan Yang and Xiaochun Cao 6904

GraphLit: Learning Text-Enriched Dynamic Character Network Representations for Literary Study
Gaspard Michel, Elena V. Epure, Romain Hennequin, Christophe Cerisara and Mirella Lapata 6920

ToM-Video: A Theory-of-Mind Video Benchmark with Spatio-Temporal Grounded Reasoning
Quanling Liu, Zizhuo Lin, Jinsheng Quan, Xin Shi, Xueqiao Zhang, Chao Zhang, Yifan Zhu, Zhihui Li and Yawei Luo 6944

VISA: Value Injection via Shielded Adaptation for Personalized LLM Alignment
Jiawei Chen, Tianzhuo Yang, Guoxi Zhang, Jiaming Ji, Yaodong Yang and Juntao Dai 6973

V-Retrver: Evidence-Driven Agentic Reasoning for Universal Multimodal Retrieval
Dongyang Chen, Chaoyang Wang, Dezhao Su, Xi Xiao, Zeyu Zhang, Jing Xiong, Qing Li, Yuzhang Shang and Shichao Kan 7000

From Agent Skills to Executable Environments: Frontier-Targeted Training for Tool-Use LLMs
Dawei Zhang, Yubin Huangfu, Kyle Yusong Yao, Boyu Wei, Zongqi Han, Likang Wu, Hongke Zhao, Ming He and Jianping Fan 7026

Emotion Experience, Expression, and Perception: Emotion Analysis on Multimodal Social Media Posts
Christopher Bagdon, Carina Silberer and Roman Klinger 7051

CoQuant: Joint Weight-Activation Subspace Projection for Mixed-Precision LLMs
Zhe Ding, Su Pan and Duowei Pan 7079

SemFlowRAG: Directed Semantic Flow from Abstraction to Evidence for Complex Reasoning
Houyuan Qin, Rong Wu, Qinyuan Qin, Botian Shi, Jingjing Qu, Yang Sun and Pinlong Cai . 7094

LLM Summaries Systematically Distort Topic Distributions
Shreeya Nallagonda, Melissa Torgbi, Harish Tayyar Madabushi and Steven Schockaert 7110

Quizzing the Translation: A Prover-Grounded Evaluation Metric for NL→FOL
Pu Suo and Ali Emami 7149

HiEvi-RAG: Hierarchical Evidence-Driven Reasoning for Long Document Understanding
Junyu Xiong, Yonghui Wang, Rongjian Gu, Chenyu Liu, Bing Yin, Wengang Zhou and Houqiang Li 7167

Charting Empirical Patterns for LLM Fine-Tuning in Scientific Multi-Discipline Learning
Lintao Wang, Zhuqiang Lu, Yilin Zhu, Kun Hu, Zhenfei Yin, Shixiang Tang, Zhiyong Wang, Wanli Ouyang and Xinzhu Ma 7182

PRISM: Private Rank-Invariant Subspace Memory for Federated Continual Fine-Tuning under Local Differential Privacy
Ziyi Wang, Liang Liu, Zhu Meng and Zhicheng Zhao 7204

eTracer: Polarity-Aware Evidence Grounding
Bohao Chu, Qianli Wang, Hendrik Damm, Shuning Zhang, Hui Wang, Andrea Papenmeier, Georg Lodde, Elisabeth Livingstone, Christoph M. Friedrich and Norbert Fuhr 7224

GSPR: Aligning LLM Safeguards as Generalizable Safety Policy Reasoners
Haoran Li, Jingru Zeng, Yulin Chen, Huihao Jing, Wenbin Hu, Hao Peng, Haochen Shi, Xi Yang, Ziqian Zeng, Sirui Han and Yangqiu Song 7254

Probing Cultural Signals in Large Language Models through Author Profiling
Valentin Lafargue, Ariel Guerra-Adames, Emmanuelle Claeys, Elouan Vuichard and Jean-Michel Loubes 7277

Mood Matters: How Syntactic Sensitivity Undermines Safety Alignment
Alina Klerings, Jannik Brinkmann, Heiner Stuckenschmidt and Simone Paolo Ponzetto 7316

DSEffi-Bench: Demystifying Large Language Models' Capability in Efficient Data Science Code Generation
Zhihao Gong, Junzhe Yu, Dong Huang, Zeyu Sun, Jie M. Zhang and Dan Hao 7337

Human Label Variation as Stable Signal: Learning Annotator-Specific Explanation Behavior via Cross-Annotator Preference Optimization
Beiduo Chen, Pingjun Hong, Ziyun Zhang, Benjamin Roth, Anna Korhonen and Barbara Plank 7373

Scaling Unsupervised Word Alignment to Documents via Structural Constraints
Michelle Wastl, Jannis Vamvas and Rico Sennrich 7419

Emergent Misalignment Is Not Magical
Mingxuan Li, Qirun Dai, Heran Wang and Chenhao Tan 7437

Fine-Grained Selective Refusal: Steering LLM Safety via Wasserstein-Informed Flow Matching
Zhiyi Liu, Haohua Du, Haoqi Sun and Haisheng Tan 7459

DOA: Training-Free Decoder-Only Attention Policy for Long-Form Simultaneous Translation with SpeechLLMs
Sara Papi and Luisa Bentivogli 7477

ViCuR: Visual Cues as Recoverable Privilege for Multimodal On-Policy Distillation
Kanghui Tian, Siyuan Liu, Ziang Yan, Sheng Xia, Shuai Dong and Yi Wang 7488

H²Table: Hierarchical Hypergraph-Enhanced Large Language Models for Complex Table Reasoning
Jia Ling, Yangfan Wang, Chen Tang, Haoming Tan, Yang Yang, Yi Guan and Jingchi Jiang . 7514

HEAL: Latent Head Alignment for Post-Training Quantization of Large Language Models
Mingzhe Yang, Sihao Lin, Ruibo Zhang, Xinyuan Zhu, Lei Chen, Weidong Chen and Xiaojun Chang 7526

ICA: Information-Aware Credit Assignment for Visually Grounded Long-Horizon Information-Seeking Agents
Cong Pang, Yujie YI, Xuyu Feng, Jiaqi Su, Zixuan Chen, Jiawei Hong, Tiankuo Yao, Nang Yuan, Jiapeng Luo, Lewei Lu and Xin Lou 7541

Bayesian Sparse Low-Rank Adaptation for Large Language Model Uncertainty Estimation
Jijie Zhang, Zhe Ren, Quan Zhang and Dandan Guo 7557

Loong: A Human-Like Long Document Translation Agent with Observe-and-Act Adaptive Context Selection
Yutong Wang, Xuebo Liu, Derek F. Wong, Zhilin Li, Rongqing Jiang, Min Zhang, Shimin Tao, Daimeng Wei and Min Zhang 7574

Is Knowledge Distillation Actually Greener? A Case Study in Machine Translation
Joseph Attieh, Timothee Mickus, Anne-Laure Ligozat, Aurélie Névéol and Jörg Tiedemann 7601

Last But Not Least: Boundary Attention CalibratiON for Multimodal KV Cache Compression
Tianhao Chen, Yuheng Wu, Kelu Yao, Xiaogang Xu, Xiaobin Hu and Dongman Lee 7619

EviRank: Structured Relevance Evidence for Multimodal Image Re-ranking
Enjun Du, Siyi Liu, Zirong Chen, Xinyu Zuo, Jinwen Luo, Ruiwen Tao, Lisheng Duan, Haijin Liang, Jin Ma, Junfu Pu and Yongqi Zhang 7640

From Sampled Outcomes to Capability Distributions: Rethinking Supervision for LLM Routing
Guannan Lai, Haoran Hu, Long Chen, Zhenguo Li and Han-Jia Ye 7667

Right Diagnoses, Decorative Reasoning: A Perturbation Audit of Medical Chain-of-Thought
Mengzhu Xu, Jifan Gao, Xia Jiang, Yaoxin Wu and Xi Long 7689

Finding Where the Buck Stops: An Automated Failure Attribution-Based Reflection Framework for Multi-Agent Collaboration
Xiaoqing Wang, Keman Huang, Bin Liang, Hongyu Li, Xiaoyong Du and Wuqiong Pan . . . 7708

When Should Models Change Their Minds? Contextual Belief Management in Large Language Models
Haoming Xu, Weihong Xu, Zongrui Li, Mengru Wang, Yunzhi Yao, Chiyu Wu, Jin Shang, Yu Gong and Shumin Deng 7729

Assessing the Impact of Language Disparity on Multilingual Linguistic Ability in Large Language Models
Zhanyu Chen and Jaap Jumelet 7752

Cognitive Honeypots: Auditing Observable Evidence of Scheming-Related Behavior in Large Language Models
Hao Wang, Wentao Chen and Deyi Xiong 7777

Aligning Cultural Behaviors in Arabic Language Models via Preferences
Junhong Liang, Daniil Orel, Saeed Almheiri, Bilal Elbouardi, Younes Samih, Abed Alhakim Freihat, Zhuohan Xie, Momina Ahsan, Muhammad Dehan Al Kautsar, Sarfraz Ahmad, Omar El Herraoui, Mohamed Anwar, Preslav Nakov and Fajri Koto 7810

CURP: Codebook-based Continuous User Representation for Personalized Generation with LLMs
Liang Wang, Xinyi Mou, Xiaoyou Liu, Xuanjing Huang and Zhongyu Wei 7839

GTS: Inference-Time Scaling of Latent Reasoning with a Learnable Gaussian Thought Sampler
Minghan Wang, Ye Bai, Thuy-Trang Vu, Ehsan Shareghi and Gholamreza Haffari 7861

Clustering-Based Balanced Sampling and Allocation with Data Parallelism for High-Performance Fine-Tuning
Hyunjin Kim, Youngeun Nam, Jaemin Han, Wonhyeok Choi and Jae-Gil Lee 7885

Auditing MCQA Benchmarks through Probability Landscapes
Minsoo Song and Chanjun Park 7904

GraSP-VL: Length as a Semantic Granularity Interface for Vision-Language Representations
Lizesheng, Pan Chengchang and Honggang Qi 7919

Directional Structurality: A Lightweight Proxy for Reasoning Data Quality
Ze Chen, Kaixuan Liu, Haomai Zhang, Jiaxuan Zou, Zhuo Chen and Xiaotian Tan 7936

Reward-Decomposed Reinforcement Learning for Immersive Video Role-Playing
Miao Wang, Yuling Shi, Yijiang Li, Yeheng Chen, Xiaodong Gu, Bin Li, Bo Gao, Wang Jun, Zengxin Han, Jingtong wu and Yaduan Ruan 7958

FLUID-V: Elastic-Horizon Diffusion Decoding for Vision-Language Models
Xiangyu Ma, Teng Xiao, Zuchao Li, Ping Wang and Hai Zhao 7990

Thinking in Different Spaces: Domain-Specific Latent Geometry Survives Cross-Architecture Translation
Marcus Armstrong, Navid Ayoobi and Arjun Mukherjee 8004

Who Speaks for the Pruned? Visual Token Pruning as Coverage Optimization
Qingchan Zhu, Weihang You, Hanqi Jiang, Changdi Yang, Tianming Liu and Geng Yuan . . 8018

From Task Success to Productive Success: Evaluating Human-AI Collaboration by Quality and Cost
Saki Imai, Mert Inan and Malihe Alikhani 8038

Not All or None: Dynamic Construction of Target-aware Memory Graph for Conversational Stance Detection
Yifan Xiang, Bin Liang, Yuqi Huang, Ruifeng Xu and Kam-Fai Wong 8057

Argus-Retriever: Vision-LLM Late-Interaction Retrieval with Region-Aware Query-Conditioned MoE for Visual Document Retrieval
Abdelrahman Abdallah, Mahmoud Abdalla, Mohammed Ali and Adam Jatowt 8075

EvoSkill Injection: Red-Teaming Autonomous Skill Generation and Evolution in Self-Evolving Agents
Doyun Kim, Chanwoo Kim, Sugyeong Eo, Yeo Chan Yoon and Chanjun Park 8095

Test-Time Scaling for Scientific Equation Discovery
Haowei Lin, Hubert Lim, Xiangyu Wang, Letian Huang and Di He 8122

PersonaDual: Balancing Personalization and Objectivity via Adaptive Reasoning
Xiaoyou Liu, Xinyi Mou, Shengbin Yue, Liang Wang, Yuqing Wang, Qiexiang Wang, Tianrui Qin and Zhongyu Wei 8144

Usage-Modulated Sentiment Representations in Large Language Models
Hongfei Du, Jiacheng Shi, Yanfu Zhang, Gang Zhou and Ashley Gao 8162

Provenance Before Prose: Claim-Locked Reporting for Statistical Text Generation
Xiao Fan, Jingyuan Li, Hongbin Guo, Yubo Han and Yi Zhang 8180

Grounding Memory Summarization in Utility Intent
Zhenyu Lei, Mingjia Shi, Xingbo Fu, Haoyu He, Qi R. Wang and Jundong Li 8195

ReGraP-LLaVA: Reasoning enabled Graph-based Personalized Large Language and Vision Assistant
Yifan Xiang, Zhenxi Zhang, Bin Li, Yixuan Weng, Bo Gao, Shoujun Zhou, Yangfan He, Yilin Yuan and Keqin Li 8210

Gradient Starvation in Low-Rank Language Model Pretraining
Victor Adelakun Omolaoye and Gerard de Melo 8249

OVLAFND: Open-Vocabulary Logic-Aware Fake News Detection
Xiaoya Fang and Lin Chen 8267

Trait Collapse in Text-to-Image Generation: Evaluating and Mitigating Personality Inconsistency in Visual Storytelling
Yuanbo Tong, Haoyuan Li, Jibiao Jiang, Zirui Wang, Yuchen Li and Jiamou Liu 8281

More Computational Resources Do Not Ensure Higher Scholarly Impact: Evidence from Leading NLP Conference Papers
Shuai Chen, Tong Bao, Jitong Peng and Chengzhi Zhang 8307

Global Consistency Repair for Natural-Language Claim Sets
Paul He, Elke Kirschbaum and Shiva Kasiviswanathan 8339

FAVoR: Measuring and Mitigating Author-Style Homogenization in Federated Personalized Generation
Lu Han, Jingyao Zhang, Katy Ilonka Gero and Nguyen H. Tran 8359

OUTLETS: Output-Length Prediction from Speculative Decoding Backbones
Weihuang Wen, Yingying Liu, Yichuan Liu, Wenqi Zeng, Li Zhou, Chumin Sun, Jie Sun and Tianshu Yu 8385

StruProKGR: A Structural and Probabilistic Framework for Sparse Knowledge Graph Reasoning
Yucan Guo, Saiping Guan, Miao Su, Jiyao Wei, Xiaolong Jin, Jiafeng Guo and Xueqi Cheng 8402

When Search Becomes Memory: Accelerating Robot Design Discovery with Self-Evolving Skills
Yunfei Wang, Xiaohao Xu, Yang Li and Xiaonan Huang 8417

Agent-as-Peer-Debriefer: A Multi-Agent Framework with Perspective-Based Refinement for Qualitative Analysis
Zhimin Lin, Kun Cheng, Zhiyao Shu, Junhua Fang, Juntao Li, Fan Bai and Jie Gao 8437

Towards Detecting AI-Assisted Responses in Online Surveys
Qizhou Wang, Bogdan Mamaev and Christopher Leckie 8466

ChainSWE: Benchmarking Coding Agents on Multi-Bug Software Maintenance
Qirui Jin, Lingching Tung, Kenan Li, Qiyang Shi, Yushi She, Huanzhong Jia, Harrison Zhao, Kejing Xia, Zhenbang Du, Jiaxin Pei, Zhenyu Zhang, Zhen Qi, Yuyan Duan, Wenke Lee and Zijian Jin 8487

TATK: Triple-Aware Top-K Learning with Knowledge-Grounded Verification for LLM-based Sequential Recommendation
Yuchen Guan, Jiaye Liu, Yifei Han, Zhenxi Zhang, Yixuan Weng and Bin Li 8506

Cross-Decoder Retargeting of Soft Compressor for Long-Context Compression
Yanhan Ye, Zhijie Nie and Richong Zhang 8529

MineExplorer: Evaluating Open-World Exploration of MLLM Agents in Minecraft
Tianjie Ju, Yueqing Sun, Zheng Wu, Wei Zhang, Yaqi Huo, Xi Su, Xierui Song, Xunliang Cai, Gongshen Liu and Zhuosheng Zhang 8542

CATMark: A Context-Aware Thresholding Framework for Robust Cross-Task Watermarking in Large Language Models
YU Zhang, Shuliang Liu, Yang Xu and Xuming Hu 8571

Reasoning Matters: Mitigate Hallucination in Multimodal Large Reasoning Models via Reasoning-Conditioned Preference Optimization
Jiawei Kong, Hao Fang, Shunxiang Liao, Jinyu Li, Bin Chen, Hao Wu, Shu-Tao Xia and Min Zhang 8601

Doc-REFRAG: Rethinking Multimodal Document Retrieval-Augmented Generation
Ruofan Hu, Shengyang Xu, Minjie Hong, Xiaoda Yang, Sashuai Zhou, Ke Lei, Tao Jin and Zhou Zhao 8624

Tracing a Sparse Emotion-Control Circuit in LLM-Based Text-to-Speech
Hongfei Du, Jiacheng Shi, Yanfu Zhang and Ashley Gao 8645

Psychologically Potent, Computationally Invisible: LLMs Generate Social-Comparison-Eliciting Posts They Fail to Detect
Hua Zhao, Jiapei Gu and Michelle Mingyue Gu 8659

OSCAR: Optimization-Steered Agentic Planning for Composed Image Retrieval
Teng Wang, Rong Shan, Jianghao Lin, Junjie Wu, Tianyi Xu, Jianping Zhang, Wenteng Chen, Changwang Zhang, Zhaoxiang Wang, Xingyu Lou, Weinan Zhang and Jun Wang 8679

Unlocking the Spoken Style: Mechanistic Localization and Lightweight Control of Orality in LLMs
Zhigen Li, Yanmeng Wang, Jianxiang Peng, Yong Zhang, Dan Shi, Ning Cheng, Shaojun Wang and Deyi Xiong 8706

CentralKV: Compressing Long-Context LLMs via Attention-Graph Centrality
Yusen Wu, Huang Yinjun, Jia Yee Tan, Hao Li, Rongfeng Guo, Liang Fan, Jiachen Luo, Guangyuan Dong and Sike Xiang 8727

When Does Visual Generation Help Visual Understanding in Unified Multimodal Models?
Yubo Zhu, Zhehan Kan, Jingyi Yang, Miaolin Chen, Jinbo Xing, Kai Zhu, Zijian Wang, Sheng Zhong and Wei Tong 8749

ToolGate: Token-Efficient Pre-Call Control for Tool-Augmented Vision-Language Agents
Anjie Liu, Yan Song, Zhixun Chen, Ziqin Gong, Zhongwei Yu and Jun Wang 8789

PlanBench-XL: Evaluating Long-Horizon Planning of LLM Tool-Use Agents in Large-Scale Tool Ecosystems
Jiayu Liu, Qihan Lin, Cheng Qian, Rui Wang, Emre Can Acikgoz, Xiaocheng Yang, Qing Zong, Jiateng Liu, Zhenhailong Wang, Xiusi Chen, Heng Ji and Dilek Hakkani-Tür 8804

Advantageous Parameter Expansion Training Makes Better Large Language Models
Naibin Gu, Yilong Chen, Zhenyu Zhang, Peng Fu, Zheng Lin, Shuohuan Wang, Yu Sun, Hua Wu, Weiping Wang and Haifeng Wang 8852

Reasoning Confidence Is Not Policy-Invariant: Calibrating Prefix Success under Model and Budget Shift
Xin Zhang, Qiyu Wei, Chenhan Yuan and Sophia Ananiadou 8875

Counter-Stereotypic Evidence Is Not Enough: Unequal Uptake in LLMs
Jinxuan Wu, Bin Li and Xiangyang Xue 8892

DualStake: Dual-Path Confidence Calibration in Deep Research Agent
Yinuo Xu, Yuwei Liang, Jianjie Cheng, Meng Wang, Yongcan Yu, Shuo Lu and Jian Liang . 8924

LegalWorld: A Life-Cycle Interactive Environment for Legal Agents
Songhan Zuo, Shengbin Yue, Tao Chiang, Guanying LI, Yun Song, Xuanjing Huang and Zhongyu Wei 8942

LQ-Rewriter: Generative Reward Model-Guided Lazy Query Rewriting for Role-Playing Dialogue
Wei Song, Zheng Huang, Renyu Zhu, Runze Wu, Zhenya Huang, Yihuan Mao, Rongsheng Zhang, Tangjie Lv and Kai Zhang 8977

Learning How to Align: Elastic Domain Adaptation for Improving Reasoning Robustness of Large Language Models
Yujie Hou, Mei Wang, Yaoyao Zhong, Guozhang Li, Qiannan Zhu and Hua Huang 8995

SAEExplainer: Interpreting SAE Features with Activation-Guided Preference Optimization
Jingyi He, Haiyan Zhao, Ruxue Shi, Yanguang Liu, Xin Wang, Fei Sun and Mengnan Du . . 9019

Rating the Pitch, Not the Product: User Evaluations of LLMs Reflect Expectations More Than Performance
Robert Morabito, Tyler McDonald, Charitra Viswanath, Angel Hsing-Chi Hwang, Susanne Gaube, Jad Kabbara and Ali Emami 9039

VerTox: Verifiable Reward-Guided Corpus Poisoning Against Neural Ranking Models
Zhiqi Huang, Vivek Datla, Zhichao Xu, Puxuan Yu, Vivek Srikumar and Alfy Samuel 9069

Who Annotates in NLP? A Large-scale Assessment of Human Annotation Reporting between 2018 and 2025
Maria Kunilovskaya, Gagan Bhatia, Lisa Sophie Albertelli, Yanran Chen, Christian Greisinger, Lotta Kiefer, Christoph Leiter, Subhadeep Roy, Tewodros Achamaleh Bizuneh, Muhammad Arslan Manzoor, Sebastian Pohl, Yufang Hou and Steffen Eger 9085

XQDT: eXplainable and Quantitative Data-Text Alignment Metric with Feedback Signals
Kun Efimov-Zhang, Yifei Song and Claire Gardent 9121

Delegated Misalignment: How Multi-Agent Structures Amplify LLM Safety Risks
Zonghao Ying, Jiaqi Yan, Huize Luo, Quanchen Zou, Aishan Liu and Xianglong Liu 9158

HALT: Hallucination Assessment via Log-probs as Time series
Ahmad Shapiro, Karan Taneja and Ashok Goel 9174

Garbage Attention in Large Language Models: BOS Sink Heads and Sink-aware Pruning
Jaewon Sok, Jewon Yeom, Seonghyeon Park, Jeongjae Park and Taesup Kim 9191

Dual-Prompt-Driven Textual Knowledge Graph Generation With Heterogeneous Information Embedding for CTR Prediction
Xiling Feng, Han Liu, Yinghui Pan and Zhiwu Xu 9213

Summarization Is Not Dead Yet
Dongqi Liu, Chenxi Whitehouse, Zheng Zhao, Zhuchen Cao, Jian Li and Yabiao Wang 9231

ArgSum-X: A Structure-Aware Benchmark for Argument Summarization and Key Point Analysis
Zhiqiang Shi and Oana Cocarascu 9263

Steering LLM Viewpoints through Fabricated Evidence Injection
Xi Yang, Chang Liu, Zhenglin Huang, Haoran Li, Weiming Zhang, Jian Weng and Yangqiu Song 9310

Relational Linearity is a Predictor of Hallucinations
Yuetian Lu, Yihong Liu, Sebastian Gerstner, Lea Hirlimann, Jonas Rohweder and Hinrich Schuetze 9327

BiMTokenizer: Preserving Semantic-Acoustic Balance in Low-Bitrate Speech Tokenization via Bidirectional State-Space Modeling
Xin Zhang, Lin Li, Chuanbo Liu, Jianquan Liu and Kong Aik Lee 9346

Detection Is Not The End: Personalized Modeling of User Susceptibility to Misinformation
Peng Qi, Zehong Yan, Wynne Hsu and Mong-Li Lee 9365

BORA-BUDGET: Learning When to Think Under a Token Budget
Zilong Wang, Kang Pan, Xiao Zhang, Erli Meng and Shuai Li 9380

ToolSpec: Accelerating Tool Calling via Schema-Aware and Retrieval-Augmented Speculative Decoding
Heming Xia, Yongqi Li, Cunxiao Du, Mingbo Song and Wenjie Li 9401

TokenCliff: Benchmarking Output-Token Budget Sensitivity in Multi-Turn LLM Agents
Bairui Zhang, Weixuan Liu, Defan Xue and Yongqi Zhang 9415

Let Them Steal: Trapping Large Language Model Extraction Attacks with Knowledge Honeypot
Yuyang Dai and Yushun Dong 9428

PromptEmbedder: Efficient and Transferable Text Embedding via Dual-LLM Soft Prompting
Yu-Che Tsai, Kuan-Yu Chen, Yuan-Hao Chen, Yu-Han Chang, Ching-Yu Tsai, Yu-Hsiang Chuang and Shou-De Lin 9450

DuoPRAG: Rank-Disentangled and Gated Knowledge-Reasoning Fusion
Nakhyeon Choi, Cheoneum Park and Seung-Hoon Na 9465

Enhancing LLMs in Predictive Political QA with Semi-Structured Data
Yinan Liu, Zihan Zhou, Zichun Jin, Xinyu Wang, Bin Wang and Xiaochun Yang 9490

Who Am I? History-Aware Profiles for Student Simulation in Tutoring Dialogues
Zhangqi Duan, Shuyan Huang, Alexander Scarlatos, Jaewook Lee, Simon Woodhead and Andrew Lan 9504

Not All Prompts Are Equal: Exploration-Guided Prompt Scaffolding for Multimodal Reinforcement Post-Training
Yuanhao Yue, Qianli Ma, Chengyu Wang, Haoting Wang, Lei Shen and Jun Huang 9522

Ada-DeVL: Adaptive Decoupling of Visual and Logical Reasoning Pathways for Efficient MLLMs
Haoyuan Fan, Jie Zhao, Yufeng Shi, Wengang Zhou and Houqiang Li 9540

What Are We Measuring in NLG? A Meta-Analysis of Evaluation Trends 2020-2025
Jing Yang, Nils Feldhus, Salar Mohtaj, Leonhard Hennig, Qianli Wang, Eleni Metheniti, Sherzod Hakimov, Charlott Jakob, Veronika Solopova, Konrad Rieck, David Schlangen, Sebastian Möller and Vera Schmitt 9563

Beyond Safe Answers: Segment-Aware Listwise Alignment for Reasoning Safety in Large Reasoning Models
JungMin Yun, Junehyoung Kwon, Hayeong Ryu, Byeonggeuk Lim, Hoejoon Kwon and YoungBin Kim 9599

ARCA: Attribution-Regularized Causal Alignment for Task-Aware Bias Mitigation
Botao Wang, Lin Li, Kaize Shi, Jiahao Liu, Chong Lyu and Xian Zhong 9616

Editing Rules in LLMs Requires More Than a Localized Update
Yating Wang, Wenting Zhao, Yaqi Zhao, Yongshun Gong, Yilong Yin and Haoliang Sun . . . 9634

Caption-once, Frames-on-Demand: Visual-Need Routing for Budget-Aware Agentic Long Video Understanding
Weitong Cai, Hang Zhang, Yukai Huang, Yiqiao Xie, Shan Gao, Jiankang Deng, Songcen Xu, Jifei Song and Zhensong Zhang 9651

MemGuard: Persisting Verifier Signals for LLM-Agent Memory Governance
Haoyu Wang, Guangyuan Dong, Liang He, Zijing Zhang, Jiachen Luo, Chao Xue, Chuang Liu and Hao Tang 9677

When Similar Means Different: Evaluating LLMs on Arabic–Hebrew Cognates
Junhong Liang, Noor Abo Mokh and Bashar Alhafni 9709

Better Retrieval, Worse Robustness: How Multi-hop RAG Amplifies Upstream ASR Errors
Zhenghua Bao 9742

Prefilling-dLLM: Predictive Prefilling for Long-Context Inference in Diffusion Language Models
Jing Xiong, Qi Han, Shansan Gong, Yunta Hsieh, Boyuan Zheng, Chengyue Wu, Chaofan Tao, Chenyang Zhao and Ngai Wong 9757

T-Mem: Memory That Anticipates, Not Archives
Weidong Guo, Dakai Wang, Zixuan Wang, Hui Liu and Yu Xu 9775

Promptology: A Large-Scale Systematic Analysis of Prompt Elements and Contextual Factors in Argument Generation
Maximilian Maurer, Ana Lisboa Cotovio, Matteo Melis, Julia Romberg, Aldo Costa and Gabriella Lapesa 9794

ToolWeave: Structured Synthesis of Complex Multi-Turn Tool-Calling Dialogues
Dinesh Khandelwal, Gnana Prakash Punnavajhala, G P Shrivatsa Bhargav, Gaurav Pandey, Sachindra Joshi, Hima Karanam and Dinesh Raghu 9827

Deep and shallow biases in language models
An Vo, Vy Tuong Dang, Khai-Nguyen Nguyen, Emilio Villa-Cueva, Thamar Solorio, Anh Totti Nguyen and Daeyoung Kim 9862

Discourse Dependency: A Continuous Criterion for Translation Difficulty
Ahrii Kim, Chanjun Park and Seong-heum Kim 9887

From Visual Cues to Spoken Narration: Rethinking Audio Description
Akshita Gupta, Aditya Arora, Federico Tombari, Marcus Rohrbach and Anna Rohrbach 9913

Read the Paper, Write the Code: Agentic Reproduction of Social-Science Results
Benjamin Kohler, David Zollikofer, Johanna Einsiedler, Alexander Miserlis Hoyle and Elliott Ash 9939

Do MLLMs Really See It: Reinforcing Visual Attention in Multimodal LLMs
Siqu Ou, Tianrui Wan, Zhiyuan Zhao, Junyu Gao and Xuelong Li 9981

Beyond a Single Story: Meta-Reviewing Sparse and Incomplete User-generated Contents for Recommendation
Hongren Wang, Tianjun Wei, Yingpeng Du, Jie Zhang and Theng Yin Leng 9998

Post-hoc Alignment of LLM-judges to Human Judgment Distribution
Sebastian Steindl, Nikos Voskarides, Alberto Gasparin and Diego Marcheggiani 10014

Time is Not a Label: Continuous Phase Rotation for Temporal Knowledge Graphs and Agentic Memory
Weixian Waylon Li, Jiaxin Zhang, Xianan Jim Yang, Tiejun Ma and Steven Y. Guo 10038

From Profiles to Steering Vectors: Global Sparse Priors and Local Semantic Calibration for Personalized Text Generation
Liuji Chen, Zeyu Zhang, Xinyuan Zhang, Shuai Nie, Qiang Liu, Shu Wu and Liang Wang 10062

Large Language Models Systematically Favor Popular Options: Evidence and Mitigation Across MCQs
Abdelrahman Abdallah, Mohammed Ali, Bhawna Piryani, Mahmoud Abdalla and Adam Jatowt 10083

Polyglot Teachers: Evaluating Language Models for Multilingual Synthetic Data Generation
Lester James Validad Miranda, Ivan Vulić and Anna Korhonen 10116

Cross-lingual Biography Enrichment via Claim Extraction and Alignment
Yifei Song, Ziyang Chen, Emil Sayilov and Claire Gardent 10145

Unified Work Embeddings: Contrastive Learning of a Bidirectional Multi-task Ranker
Matthias De Lange, Warre Veys, Jens-Joris Decorte and Jeroen Van Hautte 10196

Peak-Then-Collapse and the Four Interface Channels of Knowledge-Graph Tool Use
Tianda Sun and Dimitar Kazakov 10214

PAPO: Stabilizing Rubric Integration Training via Decoupled Advantage Normalization
Zelin Tan, Zhouliang Yu, Bohan Lin, Zijie Geng, Hejia Geng, Yudong Zhang, Mulei Zhang, Yang Chen, Shuyue Hu, Zhenfei Yin, Chen Zhang and Lei Bai 10231

Stranger, Fan, or Peer? A Systematic Study on the Role of Interlocutor in Persona-Based Dialogue Generation
Daniela Occhipinti, Malvina Nissim and Marco Guerini 10250

Black-Box Membership Inference via Word-Level Probability Estimation
Shengjie Niu, Yeheng Ge and Jian Huang 10271

Does task decomposition improve automatic NLG evaluation?
Sebastian Steindl, Nikos Voskarides, Alberto Gasparin and Diego Marcheggiani 10288

How Many Pixels Is a Digit Worth? Place-Aware Coordinate Entropy for GUI Agent Confidence Estimation
Yunxiang Li, Xixin Wu and Helen M. Meng 10302

Co-Evolutionary Prompt Optimization with Cross-Category Transfer for Zero-Shot Anomaly Detection
Sisi Zhu, Changwei Yu, Renshuai Tao and Zhenliang Ni 10316

UrduMMLU: A Massive Multitask Benchmark for Urdu Language Understanding
Ahmer Tabassum, Sarfraz Ahmad, Hasan Iqbal, Owais Aijaz, Momina Ahsan and Preslav Nakov 10341

ADWIN: Adaptive Windows for Horizon-Aware On-Policy Distillation
Kun Liang, Chenming Tang, Clive Bai, Weijie Liu, Saiyong Yang and Yunfang Wu 10371

GRASP: Gated Regression-Aware Skill Proposer for Self-Improving LLM Agents
Johannes Moll, Jean-Philippe Corbeil, Jiazhen Pan, Martin Hadamitzky, Daniel Rueckert, Lisa C. Adams and Keno Bressem 10390

How LLMs Follow Instructions: Skillful Coordination, Not a Universal Mechanism
Elisabetta Rocchetti and Alfio Ferrara 10416

Validator-Guided Decoding for Multi-Hop Knowledge Editing
Shanbao Qiao, Xuebing Liu and Seung-Hoon Na 10433

Weakening Neurons: An Input-Output Functionality in Transformers with Outsize Influence
Sebastian Gerstner, Hilal AlQuabeh, Kentaro Inui and Hinrich Schuetze 10450

Invisible to Humans, Triggered by Agents: Stealthy Jailbreak Attacks on Mobile Vision–Language Agents
Renhua Ding, Xiao Yang, Zhengwei Fang, Jun Luo and Kun He 10495

LatentMem: Customizing Latent Memory for Multi-Agent Systems
Muxin Fu, Xiangyuan Xue, Yafu Li, Zefeng He, Siyuan Huang, Xiaoye Qu, Yu Cheng and Yang Yang 10512

A Novel Semantic Manifold Alignment Attack against Embedding-to-Embedding Obfuscation in Privacy-Preserving LLMs
Sicong Li, Lingfeng Yao, Xingke Yang, Ke Tu, Chenhao Wu, Hao Wang, Jiang Liu, Phone Lin, Xin Fu and Miao Pan 10534

TableCache: Primary Foreign Key Guided KV Cache Precomputation for Low Latency Text-to-SQL
Jinbo Su, Yuxuan Hu, Cuiping Li, Hong Chen, Jia Li, Lintao Ma and Jing Zhang 10551

SmallReason-ColBERT: An Ultra-Small Late-Interaction Retriever for Reasoning Intensive Retrieval
Abdelrahman Abdallah, Mohammed Ali and Adam Jatowt 10564

MetaRAG: Belief-Action Aligned Policy Optimization for Agentic RAG
Qiuyi Qi, Tian Liang, Jiamu Wang, Jinjian Zhang, Wei Zhou, Pengcheng Zhu, Linjian Mo, Ming Kong, Jie Liu and Qiang Zhu 10576

NestedKV: Nested Memory Routing for Long-Context KV Cache Compression
Hong Chen, Xiang Liu, Yubo Gao, Yuxuan Fan, Bo Wang, Yuanlin Chu, Yuanguo Lin and Xuming Hu 10598

Beyond Input Understanding: Diagnosing Multilingual Mathematical Reasoning with Directed Acyclic Trace Graphs
Jiaqiao Zhang, Zhoujun Li, Raoyuan Zhao, Jian Lan, Thomas Seidl, Michael A. Hedderich, Hinrich Schuetze and Yihong Liu 10615

From Rollouts to Recipes: Self-Contained Post-Training for LLMs
Yifei Li, Lingling Zhang, Muye Huang, Zihan Ma, Jiashuai Liu and Jun Liu 10640

When the Bottleneck Shifts: Diagnosing and Closing Information Gaps in Interactive Text-to-SQL
Mengyi Yan, Zhinian Ma, Yue Wang, Yaoshu Wang, Xiaoke Zhu, Zizhe Wang, Haoyi Zhou and Zhumin Chen 10655

TBLS: Transition-Based Language Similarity through Activation Perturbation
Tao Ren, Johannes Bjerva and Qiongxiu Li 10688

MedQA-MM: Shortcuts Behind Medical Visual Reasoning
Benlu Wang, Yifan Zhang, Jiaqing Yu, Chin Siang Ong, Juncheng Huang, Zhuohao Li, Zhenyu Zhang, Arman Cohan, Hong yu and Zonghai Yao 10708

Mechanism Alignment for Socially Embedded Decision Making in LLMs
Jiaxin Liu, Yuhe Wu, Yixuan Xie, Hanzhi Xiao, Zeyu ma, Wenqing GU, Zeheng Xian and Guang Zhang 10759

MMedAgentBench: A Multi-Stage Benchmark for Evaluating LLM-based Medical Agents
Lu Sun, Xiangyi Zhang, Xiangyang Zhu, Kaiyuan Ji, Zijian Chen, Qiang Hu, Zicheng Zhang, Yuan Tian and Guangtao Zhai 10780

OptiMAS: Automatically Optimize Multi-Agent System
Yuxin Cheng, Chang Liu, Hanxin Yu, Haochen Tan, Taiqiang Wu, Weiqiang Jin, Jie Ran, Kaibo Wang, Xiaoguang Li, Haoli Bai, Graziano Chesi and Ngai Wong 10794

Do VLMs Share Safety Neurons Across Modalities?
Jiaxuan Li, Jiahao Zhang, Duc Minh Vo, Huy H. Nguyen, Pride Kavumba and Koki Wataoka 10822

OCGQuant: Outlier-Companion Grouping for NVFP4 Quantization
Yishan Yao, Binjun Li, Hanling Yi, Pengyu Li, Xiaoqing Liu, Zihan Yang, Xiaotian Yu and Zhiwen Yu 10838

Benchmarking LLM-as-a-Judge for Long-Form Output Evaluation
Junjie Chen, Yuxi Dong, Haitao Li, Weihang Su, Yujia Zhou, Min Zhang, Yiqun Liu and Qingyao Ai 10856

PaperAudit-Bench: Benchmarking Error Detection in Research Papers for Critical Automated Peer Review
Songjun Tu, Yiwen Ma, Jiahao Lin, Qichao Zhang, Xiangyuan Lan, Junfeng. Li, Nan Xu, Linjing Li and Dongbin Zhao 10874

Last Step Matters: Early Uncertainty Cannot Predict Failure in Long-Horizon Agents
Zongyue Li, Chengyue Yu, Lei Zang, Chenyi Zhuang, Linjian Mo and Leilei Gan 10913

When VLMs 'Fix' Students: Identifying and Penalizing Over-Correction in the Evaluation of Multi-line Handwritten Math OCR
Jin Seong, Wencke Liermann, Minho Kim, Jong-hun Shin and Soojong Lim 10932

CamoDocs: A Poisoning Attack Against Retrieval-Augmented Language Models Using Camouflaged Documents
Jaewon Jung, Haizhong Zheng, Hongsun Jang, Jaeyong Song, Beidi Chen and Jinho Lee . . 10954

MoCA-Agent: A Market-of-Claims Code Agent for Financial and Numerical Reasoning
Abdelrahman Abdallah, AbdelRahim A. Elmadany, Sameh Al Natour, Hasan Cavusoglu, Adam Jatowt and Muhammad Abdul-Mageed 10980

Open Human Value Mining: Constrained Argument Extraction and Value Concept Induction with LLMs
Shuo Wang, Binghao Fu, Ruikang Zhang, Jun Wang and Qi Su 11010

PGPO: Prototype-Guided Progressive Obfuscation for Privacy-Preserving LLM-Enhanced Recommendation
Shengxiang Lin, Jiajie Su, Pengyang Zhou, Xiang Chen, Xiaolin Zheng, Feng Tian and Chaochao Chen 11027

Multi-Source Retrieval and Reasoning for Legal Sentencing Prediction
Junjie Chen, Haitao Li, Qilei Zhang, Zhenghua Li, Ya Zhang, Quan Zhou, Cheng Luo, Yiqun Liu, Min Zhang, Yueyue Wu, Dongsheng Guo and Qingyao Ai 11047

PathSoft: Hybrid Soft Prompting with Ranked Reasoning Paths for Knowledge Graph Question Answering
Yan Teng, Xiaoding Zhou, Jinshuo Zhang, Ying Lian, Guoqiang Chen, Hongjiao Guan, Weiyu Zhang and Wenpeng Lu 11063

MedSNIP: Building and Benchmarking Snippet-Level Granularity for Medical Fact Verification
Hasan Iqbal, Sarfraz Ahmad, Hyunjae Kim, Sihyeon Park, Junjie Liao, Qingyu Chen, Preslav Nakov and Yuxia Wang 11081

Research Design Tracking and Assessment for the Social Sciences
Marco Rovera, Sergiu Burlacu, Dominique Cappelletti, Alessio Tomelleri, Sonia Marzadro, Martina Bazzoli, Annalisa Tassi and Jessica Gagete-Miranda 11105

ContraGAND: Auditing and Repairing Gender Ambiguity Failures in LLMs with Neurosymbolic Contrastive Data Augmentation
Thomas Moerman, Eleni Gkovedarou and Janiça Hackenbuchner 11136

The Weakest Link Tells It All: Outcome-Supervised Process Reward Modeling via Learnable Credit Assignment
Tianyu Jia, Yue Fang, Hongxin Ding, Rihong Qiu, Zhibang Yang, Zhijing Wu, Junfeng Zhao, Yasha Wang and Xu Chu 11160

Reactivating Test-Time Scaling for Plane Geometry Problem Solving
Xiaoqiang Kang, Shengen Wu, Maizhen Ning, Xiaobo Jin, Kaizhu Huang, Yutao Yue, Xiaowei Huang and Qiufeng Wang 11193

When Seeing Is Not Enough: Benchmarking Interactive Visual Grounding in LVLMs
Zhengxiang Wang and Owen Rambow 11212

Measuring Data Diversity with Embeddings: A Taxonomy and Benchmark for Diversity Measures
Cantao Su, Anna Wegmann, Esther Ploeger and Dong Nguyen 11241

OmniInteract: Benchmarking Real-World Streaming Interaction for Real-Time Omnimodal Assistants
Xudong Lu, Xueying Li, Annan Wang, Yang Bo, Jinpeng Chen, Zengliang Li, Nianzu Yang, Rui Liu, Xue Yang, Jingwen Hou and Hongsheng Li 11272

Dynamic Dual-Granularity Skill Bank for Agentic RL
Songjun Tu, Chengdong Xu, Qichao Zhang, Yaocheng Zhang, Xiangyuan Lan, Linjing Li, Dong Li and Dongbin Zhao 11294

DART: Low-Latency Parallel Drafting with Continuity-Aware Tree Pruning for Speculative Decoding
Fuliang Liu, Xue Li, Ketai Zhao, Yinxi Gao, Ziyan Zhou, Zhonghui Zhang, Wanchun Dou, Sheng Zhong, Zhibin Wang and Chen Tian 11316

The Multilingual FrameNet Corpus
Beatrice Fiumanò, Nicolas Lazzari, Simone Paolo Ponzetto and Valentina Presutti 11332

OBLIQ-IR: Training a Dense Retriever for Oblique Queries
Mahmoud Abdalla, Abdelrahman Abdallah, Shaimaa Sedek and Adam Jatowt 11348

When Scaling Fails: Mitigating Audio Perception Decay of LALMs via Multi-Step Perception-Aware Reasoning
Ruixiang Mao, Xiangnan Ma, Dan Chen, Ziming Zhu, Yuan Ge, Haishu Zhao, AoKai Hao, Yifu Huo, Qing Yang, Kaiyan Chang, Xiaoqian Liu, Chenglong Wang, Weiqiao Shan, Qiaozhi He, Changliang Li, Zhengtao Yu, Tong Xiao and JingBo Zhu 11370

CSAVocoder: A Causal Spatial Audio Vocoder Towards Real-Time Spatial Audio Generation
Zhiyuan Zhu, Han Wang, Wenxiang Guo, Yu Zhang, Changhao Pan, Rui Yang and Zhou Zhao 11394

FlexiSLM: A Spoken Language Model with Dynamic and Controllable Frame Rates
Jiaqi Li, Chaoren Wang, Xiaohai Tian, Mingjie Chen, Xinyu Liang, Xu Li, Yufan Lin, Junwen Qiu, Jun Zhang, Lu Lu and Zhizheng Wu 11417

Do Compounds and Phrases Follow the Same Combinatorial Principles? What a Multi-Dimensional Probing Reveals for Chinese
Qiliang Liang, Zhong Liang, Hansi Wang and Yang Liu 11441

When Humans and LLMs Disagree: Understanding Human-LLM Disagreement in Learner Intent Annotation for Educational Chatbot Dialogues
Isabel Schorr, Lisa Bardach, Haoxiang Gao and Enkelejda Kasneci 11462

When Less is More: Understanding When Token Filtering Helps and Fails in AI-generated Text Detection
Xiaoyang Han, Lvxiaowei Xu and Ming Cai 11488

Stephanie2: Thinking, Waiting, and Making Decisions Like Humans in Step-by-Step AI Social Chat
Hao Yang, Hongyuan Lu, Dingkang Yang, Wenliang Yang, Peng Sun, Wei Zuo, Xiaochuan Zhang, Jun Xiao, Kefan He, Wai Lam, Yang Liu and Xinhua Zeng 11505

Correct Prediction, Wrong Steps? Consensus Reasoning Knowledge Graph for Robust Chain-of-Thought Synthesis
Zipeng Ling, Shuliang Liu, Seonil Son, Shenghong Fu, Yuehao Tang, Yao Wan and Xuming Hu 11523

Aggregate While Searching: Online Partial-Trajectory Aggregation for Long-Horizon LLM Search Agents
Feixue Liu, Ruoyu Chen, Songyi Cui, Yimo Yan and Yong-Hong Kuo 11555

Why and When to Use Statistical Significance in Explanation Evaluation: A Systematic Survey and Guidelines for Explainable NLP
Fariz Ikhwantri and Dusica Marijan 11571

Truth or Tribe: How In-group Favoritism Prioritize Facts in Persona Agents
Shijun Lei, Hongyu Wang, Yunji Liang, Haowen Zheng, Bin Guo and Zhiwen Yu 11596

CausalTAD: Injecting Causal Knowledge into Large Language Models for Tabular Anomaly Detection
Ruiqi Wang, Ruikang Liu, Runyu Chen, Haoxiang Suo, Zhiyi Peng, Zhuo Tang and Changjian Chen 11618

GSEM: Graph-based Self-Evolving Memory for Experience Augmented Clinical Reasoning
Han Xiao, Yuzheng Fan, Sendong Zhao, Haochun Wang and Bing Qin 11637

An Empirical Study on Preference Tuning Generalization and Diversity Under Domain Shift
Constantinos Karouzos, Xingwei Tan and Nikolaos Aletras 11661

Evidence Laundering: Deep Research Agents Mistake Correlated Sources for Independent Evidence
Ruihua Zhou, Hao Huang, Yijie Zheng, JiaTang Luo, Yunpeng Li, Yuling Liu and Dan Du 11682

Summoning the Oracle to Slay It: Mitigating Look-Ahead Bias in Financial Backtesting with Large Language Models
Weixian Waylon Li, Mengyu Wang and Tiejun Ma 11694

Every Time I Hire a Linguist, Inference Costs Go Down: On Linguistic Rules as Effective Prompt Compressors
Jianfei Ma, Zhaoxin Feng, Emmanuele Chersoni and Si Chen 11714

TopoAlign: A Framework for Aligning Code to Math via Topological Decomposition
Yupei Li, Philipp Borchert and Gerasimos Lampouras 11751

Beyond SFT-Optimality: Cold-Start Data Selection for Reinforcement Learning via Gradient Subspace Coverage
Haisheng Lv, Zikun Su, Jingbang Liu, Zilu Huang, Xinzhang Liu, Wenmin Deng, Yu Zhao, Chao Wang and Shuangyong Song 11769

LLM Agents as Computational Typologists
Changbing Yang, Christopher Hammerly, Freda Shi and Jian Zhu 11790

Personas Differ from Native-Language Generation: Language Pathways Shape LLM Interpersonal Advice
Jinhee Won and Xinlan Emily Hu 11816

Beyond the Vacuum: Combinatorial Strategy Selection for Competitor-Aware Generative Engine Optimization
Vaibhav Sourirajan, Yao Zhang, Himanshu Kumar, Sahil Wadhwa, Mann Patel and Amirfarrokh Iranitalab 11852

The Magnitude Mirage: Rethinking Confidence for Reasoning-Intensive Retrieval
Jamie Holdcroft, Abdelrahman Abdallah and Adam Jatowt 11872

CoFiE: Coarse-to-Fine Evidence Selection for Efficient Streaming Video Understanding
Jing Jiang, Yiran Ling, Ruonan Li, Dimitrios Stamoulis and Jie Liu 11893

One-Token Verification for Reasoning Correctness Estimation
Zhan Zhuang, Xiequn Wang, Zebin Chen, Feiyang Ye, Ying Wei, Kede Ma and Yu Zhang 11906

Universal Activation Verbalizer: A Unified Framework for Cross-Model Activation Explanation
Haiyan Zhao, Zirui He, Guanchu Wang, Ali Payani, Yingcong Li and Mengnan Du 11931

Bridging the Consistency Gap: A Noun Bank Augmented Framework for Low-Cost Long Text Translation
Tao Chen, Mao Zheng, Zheng Li, Zhang Zheng and Bo Lv 11960

Evaluating and Improving LLM Self-Modeling
Siqi Zeng, Andre N. Assis and Rowan Wang 11978

Stick to What You Know: A Study of Knowledge-Aligned Supervised Fine-Tuning
Arthur Becker, Jakob Kemmler, David Thulke, Christine Schäfer, Christian Dugast and Hermann Ney 12063

One Size Does Not Fit All: Token-Wise Adaptive Compression for KV Cache
Liming Lu, Kaixi Qiu, Jiayu Zhou, Jushi Kai, Haoyan Zhang, Jingzhi Wang, Huanyu Wang, Bo Jiang, Jingwen Leng, Ziwei He and Zhouhan Lin 12100

WRIT: Write-Read Intensive Trajectory Synthesis for Multi-Turn User-Facing Agents
Hengrui Gu, Xiaotian Han and Kaixiong Zhou 12115

Beyond: Better-than-Full-Precision KV Caches via Learnable Non-Uniform Quantization as Implicit Regularization
Yuhao Xie and Mingjie Lin 12135

When Less Is More: An Empirical Study of Minimal Responses in Counseling Dialogues and the Behavior of LLMs
Zhiyang Qi 12159

Plasticity without Collapse: Plasticity-Inducing Control Optimizer for Cross-Lingual Continual Pre-Training
Sumin Kim, Minjun Song, Surin Lee, S M Wahidur Rahman, Yongtae Lee, Haeung Choi and Heung-No Lee 12174

CodeTD: Topology of Attention Detects Hallucinations in Code LLMs
Daria Voronkova, Ilya Trofimov, Anton Dmitriev, Eduard Tulchinskii, Evgeny Burnaev and Serguei Barannikov 12204

EARAL: Efficient Attribution of Retrieval-Augmented Answer Generation using Logits
Lucas Albarede, Jose G Moreno, Lynda Tamine and Luce Lefeuvre 12233

HTAM: Hierarchical Transition-Attended Memory for Operator Optimization
Yining Zhang, Mingyang Yi, Chen Wang, Xuwen Xiang, Tianhe Jia, Zedong Dan, Chengqing Zong and Yue Wang 12248

TraceMIA: Interpretable Membership Inference Attack via Lexical Patterns in Reasoning Traces
Bo Zhang, Yongtang Wu, Ning Liu, Jun Zhang, Zhunchen Luo, Yunbo Cao, Xiaoying Bai, Guotong Geng and Zhong Wang 12273

MoF: Preference-Aware Mixture Modeling for Black-Box LLM Personalization
Hun Park 12290

Rethinking Prospect Theory for LLMs: Revealing the Instability of Decision-Making under Epistemic Uncertainty
Rui Wang, Qihan Lin, Jiayu Liu, Qing Zong, Tianshi Zheng, Dadi Guo, Haochen Shi, Peixuan Han, Weiqi Wang and Yangqiu Song 12310

Expanding the Solvable Frontier: Policy-Adaptive Curriculum Learning for Reasoning RL
Guocong Li, Jinjian Zhang, Ping Wang, Siyan Guo, Yukai Sun, Hao Huang, Wei Zhou, Linjian Mo, Hongxia Xu and Jian Wu 12329

TIDC: Test-Time Intervention-based Diagnosis and Calibration for Harmful Meme Detection
Zhihui Guo, Xin Man, Hao Yin, Feiyu Chen and Jie Shao 12346

WebRISE: Requirement-Induced State Evaluation for MLLM-Generated Web Artifacts
Yuxin Meng, Yuhan Suo, Junjie Wang, Yuhan Sun, Yiyao Yu, Ruixu Zhang, Ruining Hu, Yubin Wang, Shouwei Ruan, Bin Wang, Yuxiang Zhang and Yujiu Yang 12360

MR^2-Bench: Going Beyond Matching to Reasoning in Multimodal Retrieval
Junjie Zhou, Ze Liu, Lei Xiong, Jin-Ge Yao, Yueze Wang, Shitao Xiao, Chaofan Li, Huaying Yuan, Fenfen Lin, Miguel Hu Chen, Zhicheng Dou, Siqi Bao, Defu Lian, Yongping Xiong and Zheng Liu 12386

CoRect: Context-Aware Logit Contrast for Hidden State Rectification to Resolve Knowledge Conflicts
Xuhua Ma, Richong Zhang and Zhijie Nie 12414

The Blindness of Document-Level Translation Evaluation
Ahrii Kim, Vilém Zouhar, Chanjun Park and Seong-heum Kim 12433

OmniRetriever: Any-to-Any Audio-Video-Text Retrieval via Fusion-as-Teacher Distillation
Yunze Liu, Chi-Hao Wu, Enmin Zhou and Junxiao Shen 12455

Axis-LoRA: Probing and Adapting Embedding Models for Personal-Memory Retrieval
Dohoon Kim, Junho Park, Hyunwoong Bae, Seyoon Kim and Taesup Moon 12473

SRAF: Stealthy and Robust Adversarial Fingerprint for Copyright Verification of Large Language Models
Zhebo Wang, Zhenhua Xu, Li Maike, Xilong Cheng, Wenpeng Xing, Chunqiang Hu, Chen Zhi and Meng Han 12496

CAD-Mark: Context-Aware Distributional Watermarking for Robust LLM-Generated Text Detection
Zhaoheng Huang, Dacheng Wen, Yutao Zhu, Zhicheng Dou, Ji-Rong Wen and Fangzhao Wu 12516

Know What You should not Know: Dynamic Attention Gating for Knowledge-Bounded Role-Playing
Zongsheng Wang, Bowen Wu, Qun yu, Dengzhengwei, Qin Lin, Xu Chen and Baoxun Wang 12530

RAGSentinel: Certifiable Geometric Consensus for Robust Retrieval-Augmented Generation
Yueyang Quan, Anjun Gao, Yufei Xia, Minghong Fang and Zhuqing Liu 12549

The "10th Juror": Open-Set Standpoint Screening for Bureaucratic Bias Detection
Yuchen Miao, Zijun Wang, Chang Han, Yurui Shi, Mingtai Zhang and Siyang xu 12576

Which one is banana man? Evaluating vision–language models in multi-turn pragmatic interpretation
Alvin Wei Ming Tan, Ben Prystawski and Veronica Boyce 12595

LLM Abstention Can Be a Prompt Artifact, in Addition to Genuine Uncertainty
Zipeng Ling, Shuliang Liu, Yuehao Tang, Junqi Yang, Shenghong Fu, Seonil Son, Chen Huang, Kejia Huang, Yao Wan, Zhichao Hou and Xuming Hu 12612

MemeCULT-1K: Benchmarking South Asian Cultural Context and Humor Understanding of Multimodal Models
Tawsif Tashwar Dipto, Mehedi Ahamed, Radib Bin Kabir, Mueeze Al Mushabbir, Mohammed Saidul Islam, Mir Rayat Imtiaz Hossain, Md Tahmid Rahman Laskar and Sabbir Ahmed 12634

Target-Language Generation in Multilingual Models: Activation Steering and Optimal Control
James A. Michaelov, Carmen Amo Alonso, Tyler A. Chang and Roger P. Levy 12644

Illuminating the Path: Reachability-Aware Supervision Curation for LLM-Based Recommendation
Donghee Han, Jiwon Jeong, Hwanjun Song and Mun Yong Yi 12665

MemoryCard: Topic-Aware Multi-Modal Clue Compression for Long-Video Question Answering
Qing Yang, Pengcheng Huang, Xinze Li, Zhenghao Liu, Yukun Yan, Yu Gu, Ge Yu, Gang Li and Maosong Sun 12687

Emotion Understanding in Streaming Video with Trajectory-Aware Reliability
Qingsong Wang, Qigong Lei, Zitong Wang, Bohan Yu, Zhiang Dong, Jian Liu, Weiqiang Wang, Chang Yao and Jingyuan Chen 12710

SafeSteer: Localized On-Policy Distillation for Efficient Safety Alignment
Hao Li, Jingkun An, Zijun Song, Pengyu Zhu, Rui Li, Hao Wang, Wendi Feng, Yesheng Liu, Lijun Li, Jin-Ge Yao and Lei Sha 12729

MDM-Prime-v2: Binary Encoding and Index Shuffling Enable Scaling of Diffusion Language Models
Chen-Hao Chao, Wei-Fang Sun, Junwei Quan, Chun-Yi Lee and Rahul G Krishnan 12748

Accelerating Neural Architecture Search with Code Semantics
Zhen Liu, Wanqi Zhou, Shuanghao Bai, Wei Song, He Zhang, Yuhan Liu and Jingwen Fu . 12778

Finding What Matters: Anchoring Context Knowledge with Evolving Indices for Iterative Retrieval
Mingyan Wu, Zhenghao Liu, Xinze Li, Yuqing Lan, Yukun Yan, Shuo Wang, Cheng Yang, Minghe Yu, Zheni Zeng and Maosong Sun 12796

Why Are Some Emotions Harder for LLMs? Uncovering the Causal Mechanisms of Emotion Inference via Sparse Autoencoders
Bangzhao Shu, Arinjay Singh and Mai ElSherief 12821

Understanding and Mitigating Cross-lingual Privacy Leakage via Language-specific and Universal Privacy Neurons
Wenshuo Dong, Qingsong Yang, Shu Yang, Lijie Hu, Meng Ding, Wanyu Lin, Tianhang Zheng and Di Wang 12842

Paint It, Black: Evaluating LLMs for De-Identification of Multimodal Clinical Documents
Francesco Zangrillo, Alessio Cocchieri, Giacomo Frisoni, Davide Freddi, Luca Ragazzi, Alberto Aguzzi, Diego Raimondo, Renato Seracchioli and Gianluca Moro 12865

What Can Reviews Reliably Tell Us About Books? A Per-Attribute Audit and Tiered Annotated Dataset Across 1.7M Titles
Samsun Knight, Yakov Bart, Ethan Tang and Jory Gutman 12903

SGTR: Learning Style-Generalizable Trajectory Reasoning for Zero-Shot Text Style Transfer
Yuan Gao, Yuwei He, Yashuo Luo, Junlin Hu and Jun Yang 12935

SkillMe: Evolving Agent Skills for Recurring Personal Workflows
Xiuqi Lu, Jie Zhang, Yuanfu Liu, Lewen Yan, Yutong Zhi and Jing Shao 12951

MemoryForge: Synthesize Lifelong Memory for Human-Like LLM Agents
Bohan Tang and Steven Y. Guo 12974

From Rubrics to Reliable Scores: Evidence-Grounded Text Evaluation with LLM Judges
Yihan Hong, Huaiyuan Yao, Bolin Shen, Wanpeng Xu, Hua Wei and Yushun Dong 13001

Does Fine-Tuning Undo Activation Steering? Behavioural Recovery Without Weight-Edit Reversal
Philipp E. Glass, Allan Tucker, Yongmin Li and Alina Miron 13022

Natural Language to First-Order Logic LLM-based Autoformalization
Andrea Brunello, Cristian Curaba, Luca Geatti, Michele Mignani, Angelo Montanari and Nicola Saccomanno 13036

Mimir: Large-Scale Multilingual Concept Modeling
Elio Musacchio, Lucia Siciliani and Pierpaolo Basile 13051

P-Search: A Parallel Plan–Act–Verify Framework with Decoupled GRPO for Deep Research
Pin Zheng, Yibo Xue, Ming Wang, Yukun Kang, Haonan Zheng, Chengyu Wang, Jian Peng, Jun Huang and Changliang Xu 13074

Dual-View Training for Instruction-Following Information Retrieval
Qingcheng Zeng, Puxuan Yu, Aman Mehta, Fuheng Zhao and Rajhans Samdani 13099

STEB: Style Text Embedding Benchmark
Rafael Alberto Rivera Soto, Anna Wegmann and Cristina Aggazzotti 13109

Overtone: Fine-Grained Audio Reasoning via Diagnostic-Driven Delayed Acoustic Fusion
Zhuoyan Mi, Kangle Sun, Yikai Yang, Yujie Fu and Yuehai Wang 13146

SpinHash: An O(1) Encoder for Any Vocabulary
Jia Xu, Jinhua Wang and Constantin F. Aliferis 13165

LLM-Powered Holistic Enhancement for Text-Attributed Graphs
Xiaotong Zhang, Haoliang Ming, Ze Wang and Han Liu 13181

Fixing FOLIO and MALLS: Verified Annotations and an LLM-assisted Framework to Focus Human Relabeling
Andrea Brunello, Cristian Curaba, Luca Geatti, Michele Mignani, Angelo Montanari and Nicola Saccomanno 13195

EVAR: Evidence-Validated Hypothesis Admission for Budget-Aware Narrative Reasoning
Peilin Liu, Zhiquan Ji and Jinglong Ping 13221

Beyond CTC Spikes: LLM-Driven Gloss-Temporal Supervision for Continuous Sign Language Recognition
Junfeng Zhu, Cheng Jin and Yuchun Fang 13237

Routing-Free Mixture of Experts
Yilun Liu, Jinru Han, Sikuan Yan, Volker Tresp and Yunpu Ma 13253

Chartographer: Counterfactual Chart Generation for Evaluating Vision-Language Models
Yifan Jiang, Dae Yon Hwang, Jesse C. Cresswell and Freda Shi 13272

SABER: Stability-Aware Early Exit for LLM Reasoning via Adversarial Branch Probing
Wanli Cheng, Haiya Xiang, Juntao Li, Hongling Wang and Wenliang Chen 13294

When Metadata Remembers: Ordered Provenance Enables Document-Level Embedding Inversion
Liwen Zheng, Qing Li, Qingsong Zou and Yong Jiang 13314

When LLMs Forget What They Believe: Characterizing State Integrity Collapse in Long-Horizon Dialogues
Yancheng Zhu and Yifan Yuan 13334

LongDS-Bench: On the Failure of Long-Horizon Agentic Data Analysis
Kewei Xu, Xiaoben Lu, Shuofei Qiao, Zihan Ding, Haoming Xu, Lei Liang and Ningyu Zhang 13351

AgentSpec: Speculative Decoding for Batch Inference of LLM Agents
Xin Wang, Ziming Miao, Yi Zhu, Hui Shen, Zhongwei Wan, Fan Yang and Mi Zhang 13384

MeetingToM: Evaluating Multimodal LLMs on Theory-of-Mind Reasoning in Multi-Party Meetings
Ziyi Wang, Yuhang Wu, Dongxu Piao, Xingyu Liu, Tianhui Zhou and Miao Liu 13396

Investigating Linear Probe Robustness to Linguistic Register, Medical Specialty, and Corpus Shifts in Medical QA
Nishant Mishra, Ameen Abu-Hanna and Iacer Calixto 13423

From Prompts to Trees: Effective LLM-Guided Tree Generation for Few-Shot Tabular Classification
Yue Qiu, Zekang Du, Yiqun Diao, Bingsheng He and Qinbin Li 13451

SciLENS: RL-Driven Autonomous Agents for Scientific Localized Evidence Navigation and Synthesis
Leqi Zheng, Jinbo Su, Yuying Li, Chaokun Wang, Weiping Wang, Haitao Li, Jiajun Zhang, Shannan Yan, Zhaolu Kang, Rong Fu, Jie Wu, Fang Niu and Hang Zhang 13482

All for 1-Bit: Towards Genuine 1-Bit Post-Training Quantization for LLMs
Zhixiong Zhao, Zukang Xu, Guangyu Sun, Lifeng Liu and Dawei Yang 13527

Re2A: Situated Conversational Recommendation via Rubric-based Preference Reasoning and Alignment
Dongding Lin, Jian Wang, Xiaoyan Zhao and Wenjie Li 13559

Compositional Machine Design as Program Synthesis with LLMs
Wenqian Zhang, Yangyi Huang, Weiyang Liu and Zhen Liu 13586

JustifIT-QA: Evidence-Grounded Legal Question Answering in Italian
Niko Dalla Noce, Davide Bacciu, Sina Farhang Doust and Lorenzo De Mattei 13661

Gradients Know What Outcomes Don't: Unlocking Reinforcement Learning for LLM Reasoning with Gradient-Aligned Rewards
Leqi Zheng, Jinbo Su, Fang Niu, Chaokun Wang, Weiping Wang, Jiajun Zhang, Shannan Yan, Jie Wu, Zhaolu Kang, Rong Fu and Hang Zhang 13686

OntologyBench: Can Dense Retrieval Satisfy Structured Biomedical Constraints?
Xiao Yu Cindy Zhang, Wyeth Wasserman and Jian Zhu 13707

Why We Need Speech to Evaluate Speech Translation
Maike Züfle, Danni Liu, Vilém Zouhar and Jan Niehues 13726

ReactBench: A Benchmark for Topological Reasoning in MLLMs on Chemical Reaction Diagrams
Qiang Xu, Shengyuan Bai, Yu Wang, He Cao, Leqingchen, Yuanyuan Liu, Bin Feng, Zijing Liu and Yu Li 13744

sEMG Typing Transcription with Conformer Encoding and Beam-Aware Large Language Model Correction
Yuanxi Wang, Lun Shu and Yasuharu Koike 13763

Beyond Next-Observation Prediction: Agent-Authored World Modeling for Sequential Decision Making
Guangfeng Cai, Kaibing Yang, Shuo He, Yu Li, Shengtian Yang, Jiaqi Lv and Lei Feng . . . 13781

Real Images, Worse Judgments: Evaluating Vision-Language Models on Concreteness and Imagery
Yifan Jiang, Ruoxi Ning, Sheng Yao and Freda Shi 13801

InfoDensity: Rewarding Information-Dense Traces for Efficient Reasoning
Chengwei Wei, Jung-jae Kim, Longyin Zhang, Shengkai Chen and Nancy F. Chen 13829

DeskCraft: Benchmarking Desktop Agents on Professional Workflows and Human-in-the-Loop Collaboration
Wenkai Wang, Tao Xiong, Jingchen Ni, Yunpeng Bao, Xiyun Li, Tianqi Liu, Hongcan Guo, Zilong Huang and Shengyu Zhang 13844

SafeGEO: Understanding Generative Engine Optimization Risks in Recommendation Agents
Qianfeng Wen, Yifan Simon Liu, Xin Liu, Difan Jiao, Blair Yang, Junda Wu and Zhenwei Tang 13887

AWARe: Mitigating Catastrophic Forgetting via Activation-Weighted Adaptive REtention
Juncheng Liao, Jinfan Lv, Guoming Wang, Juepeng Zheng, Ling Xiao and Siliang Tang . . 13928

REAT: A Reflective Experience-Augmented Tutoring Framework for Multi-turn Mathematical Instruction
Jianheng Zhou, Chaoli Zhang, Xingjun Wei, Xinliang Zhou, Giancarlo Fortino, Xing Fan, Yanfeng Wang, Qingsong Wen and Haoyang Li 13945

InSight: A Benchmark for Agentic Claim Verification in Interactive Visualizations
Maeve Hutchinson, Syed Mahbubul Huq, Mohammad Albinhassan, Radu Jianu, Aidan Slingsby and Pranava Madhyastha 13968

Decomposing Visual Histories with Vision-Language Agents: Hierarchical Temporal Guidance for Compositional Image Generation
Lujundong Li, Mingxu Zhang, Songning Lai, Jiayu Yang, Yuxuan Fan, Ziwei Xie, Feng Liu, Changwang Zhang, Jun Wang, Dazhong Shen and Ying Sun 13989

MathDebugger: A Benchmark for Type-Aware Error Detection in Synthetic Math Data
Hao Liang, Meiyi Qiang, Yuying Li, Zefeng He, Xiaochen Ma, Ruichuan An, Yongzhen Guo, Zhengzhou Zhu, Bin Cui and Wentao Zhang 14001

Cooperate to Compete: Strategic Coordination in Multi-Agent Conquest
Abigail O'Neill, Alan Zhu, Mihran Miroyan, Narges Norouzi and Joseph E. Gonzalez 14019

Does Finetuning with Scientific Data Increase Hallucinations? A Multi-domain Factuality Evaluation of LLMs
Raia Abu Ahmad, Nikolas Rauscher, Ekaterina Borisova, Fabio Barth, Georg Rehm and Sebastian Möller 14040

Cognitive-Level Prefix-MoE Reward Modeling for Curriculum Alignment of Medical Language Models
Hao Wang, Wendi Feng, Chun Kang, Yanting Wang, Rui Li and Lei Sha 14065

WebWorld: The Browser as a World Model for Self-Improving Web Code
Jiajun Wu, Jian Yang, Yaxin Du, Wei Zhang, Haowen Wang, Junhang Cheng, Yuxuan Zhang, Tuney Zheng, Xianglong Liu and Ming Zhou 14080

IRD: Steering Reasoning Models with Intent, Risk, and Decision
Kang Zhou, Guanghao Zhou, Cen Chen, Zhaoyang Liu, Ming Gao and Yaliang Li 14103

The Mirage of Explainability: A Survey on Chain-of-Thought Faithfulness in Large Language Models
Xi Jin, LeChen Qi, Houfeng Wang and Liangming Pan 14122

Divide-and-Adapt: Fine-Tuning VLMs via Model Division and Routing
Hailing Wang, Neil Rodrigues, Xianglong Wang, Wei Li, Lin Chen and Yun Fu 14137

LLaDA-Rec: Discrete Diffusion for Parallel Semantic ID Generation in Generative Recommendation
Teng Shi, Chenglei Shen, Weijie Yu, Shen Nie, Chongxuan Li, Xiao Zhang, Ming He, Jianping Fan and Jun Xu 14152

Step-Wise Refusal Dynamics in Autoregressive and Diffusion Language Models
Eliron Rahimi, Elad Hirshel, Amit LeVi, Rom Himelstein, Avi Mendelson and Chaim Baskin 14166

FuzhouBench: A Benchmark for Eastern Min Chinese with Tone Sandhi Annotations
Xiaocan Li and Haoyu Lin 14194

Beyond Good Intentions: When Does the Framing of Multilingual and Low-Resource NLP Research Become a Caricature?
Nedjma Ousidhoum, Noopur Zambare and Mohamed Abdalla 14209

Fair Like Us? Auditing LLM Alignment in Resource Allocation
Qishen Han, Hadi Hosseini, Joshua Kavner, Samarth Khanna, Sujoy Sikdar and Lirong Xia14223

Advancing ESG Intelligence: An Expert-level Agent and Comprehensive Benchmark for Sustainable Finance
Yilei Zhao, Wentao Zhang, Lei Xiao, Yandan Zheng, Mengpu Liu and Wei Yang Bryan Lim14256

SemBridge: Language Transfer in Sparse Encoders via Multilingual Semantic Bridge
Seongtae Hong, Youngjoon Jang, Jia-Huei Ju, Hyeonseok Moon and Heuiseok Lim 14291

Length-Adaptive Decoding for Masked Diffusion Machine Translation
Yan Zhan, Mengkai Hou, Wanting Zhang and Zhijun Gao 14306

EDGE: Experience-Distillation for Guided Exploration in Agentic Reinforcement Learning
Can Xie, Yuyi Zhou, Wen Yang, Ziyi Zhang, Siyao Song, Yingzhuo Deng, Shuo Ren and Jiajun Zhang 14328

Measuring the Behavioral Fidelity of Long-Horizon Human Activity Simulations
Yi Fei Cheng, Fan Yang, Iremsu Bas, Koichiro Niinuma, Narishige Abe and David Lindlbauer 14351

SkillBrew: Multi-Objective Curation of Skill Banks for LLM Agents
Wentao Hu, Zhendong Chu, Yiming Zhang, Junda Wu, Ming Jin, Xiangyu Zhao, Yilei Shao, Yanfeng Wang and Qingsong Wen 14373

POLARIS: Guiding Small Models to Write Long Stories
Rishanth Rajendhran, Jenna Russell, Mohit Iyyer and John Frederick Wieting 14392

Making Clinical Language Models Auditable: Concept-Guided Fine-Tuning for Robust Prediction
Jin Mu and Guanhua Chen 14429

When LLMs Benchmark Themselves: Deconstructing Self-Bias in Automated Evaluation
Wenda Xu, Sweta Agrawal, Vilém Zouhar, Markus Freitag and Daniel Deutsch 14447

Recoverability-Controlled Safety Miscalibration in Large Language Models
Yuwei Wang, Xingwang Liang, Yu Xie, Haotian Chi and Haijun Geng 14463

Agreement Is Not Truth: Characterizing and Controlling Consistent-but-Wrong Reasoning in LLMs
Minje Kim, Keeheon Lee and Chi-Guhn Lee 14482

PlannerForge: LLM Agents for Scenario-Based Testing of Motion Planners in Autonomous Driving
Yuan Gao, Sebastian Müller, Mattia Piccinini, Marc Kaufeld, Yuchen Zhang, Finn Rasmus Schäfer, Qunying Song and Johannes Betz 14522

PRISM: Self-Pruning Intrinsic Selection Method for Training-Free Multimodal Data Selection
Jinhe Bi, Aniri, Zengjie Jin, Yifan Wang, Danqi Yan, Wenke Huang, Xiaowen Ma, Sikuan Yan, Artur Hecker, Mang Ye, Xun Xiao, Hinrich Schuetze, Volker Tresp and Yunpu Ma 14557

FinDeepForecast : A Live Multi-Agent System for Benchmarking Deep Research Agents in Financial Forecasting
Xiangyu Li, Xuan Yao, Guohao Qi, Fengbin Zhu, Kelvin J.L. Koa, Xiang Yao Ng, Ziyang Liu, Ni Xingyu, Liu Chang, Yonghui Yang, Yang Zhang, Wenjie Wang, Fuli Feng, Chao Wang, Huanbo Luan, Xiaofen Xing, Xiangmin Xu, Tat-Seng Chua and Ke-Wei Huang 14581

CausalChapter: Improving Long-Form Video Chaptering with Interventional Dependency Modeling
Xinran Duan, Guozhang Li, Yaoyao Zhong, Mei Wang, Lizhi Wang and Hua Huang 14616

Mitigating Overthinking via Interleaved System-1/2 Reasoning in Large Language Models
Han Ding, Wei Chen and Fuzhen Zhuang 14633

World Models for Policy Refinement in StarCraft II
Yixin Zhang, Ziyi Wang, Yiming Rong, Haoxi Wang, Jinling Jiang, Shuang Xu, Haoran Wu, Shiyu Zhou and Bo XU 14657

AgentProv: Auditing Agentic LLM API Providers via Tool-use Policy Probes
Xun Wang, Bihe Zhao, Michael Backes, Franziska Boenisch and Adam Dziedzic 14689

TIDE: Trajectory-based Diagnostic Evaluation of Test-Time Improvement in LLM Agents
Hang Yan, Xinyu Che, Fangzhi Xu, Qiushi Sun, Zichen Ding, Kanzhi Cheng, Jian Zhang, Tao Qin, Jun Liu and Qika Lin 14705

Dockerless: Environment-Free Program Verifier for Coding Agents
Wenhao Zeng, Yuling Shi, Xiaodong Gu, Chao Hu, Chaofan Wang, Yuhao Cui, Hongting Zhou, Mengnan Qi, Jianqiao Wangni, Zhaojian Yu, Shuzheng Gao, Kai Cai and Shilin He 14730

Self-Correction Can Amplify Hallucinations: Fact-Level Repair with Graph-Based Evidence Routing in Multimodal Generation
Kaixiang Zhao, Tianrun Yu, Shawn Huang, Porter Jenkins, Yushun Dong and Amanda Hughes 14747

SciFigBench: A Benchmark for Scientific Method Figure Understanding
Lijia Wu, Zheng Li, Yuxuan Liu, Jingwen Zhang, Penghai Zhao and Xiang Li 14776

Subspace Alignment for Vision-Language Model Test-time Adaptation
Zhichen Zeng, Wenxuan Bao, Xiao Lin, Ruizhong Qiu, Tianxin Wei, Xuying Ning, Yuchen Yan, Chen Luo, Monica Xiao Cheng, Jingrui He and Hanghang Tong 14801

GPAgentBench-2K: Benchmarking Large Language Model Agents in Complex Clinical Action Space
Boqi Chen, Xudong Liu, Yunke Ao, Heejin Do and Jianing Qiu 14824

Cited but Not Causal: A Counterfactual Audit of Refusal Explanations in Aligned Language Models
Minje Kim, Keeheon Lee and Chi-Guhn Lee 14854

Selective State-Space Adaptation and Retrieval for Language Model Reasoning
Atahan Dokme and Larry Heck 14878

Graph Evidence Is Not Enough: Diagnosing Native Decoder Use in Graph-Augmented LLMs
Xiaoyu Guo, Pengcheng Chen, YU Jiong, Yi Lu, Yaohua Wang and Ziyang Li 14900

SieveServe: Adaptive Visual Token Pruning for Reliable and Efficient Multimodal LLM Inference
Xiao Li, Xiao Shi, Cheng Zhang, Sa Wang, Zhining Zhang and Xiaofang Zhao 14918

Think, Act, Build: An Agentic Framework with Vision Language Models for Zero-Shot 3D Visual Grounding
Haibo Wang, Zihao Lin, Zhiyang Xu and Lifu Huang 14935

Revisiting Greedy Decoding for Visual Question Answering: A Calibration Perspective
Boqi Chen, Xudong Liu, Yunke Ao and Jianing Qiu 14954

Self-Supervised On-Policy Distillation for Omni-Modal Video Captioning
Zinuo Li, Yongxin Guo, Xin Li, Hao Tan, Zhao Zhong, Liefeng Bo, Mohammed Bennamoun, Farid Boussaid and Qiuhong Ke 14974

MiST: Mid-Training LLMs for Cybersecurity
Oded Ovadia, Elad Ben Zaken, Elad Guttman and Orly Moreno Kadosh 14996

SpatialGuard: Harness-Guided Verifiable Spatial Reasoning for Text-to-Image Generation
Ziyun Qian, Zizhi Chen, Yizhou Liu, Mingyang Sun, Dingkang Yang and Lihua Zhang . . . 15022

Beyond Decodability: Reconstructing Language Model Representations with an Encoding Probe
Gaofei Shen, Martijn Bentum, Tomas Ostar Lentz, Afra Alishahi and Grzegorz Chrupała . 15033

CogniDir: Combating Cognitive Malicious Comments via Adaptive Distributional Learning for Robust Fake News Detection
Zhao Tong, Chunlin Gong, Yimeng Gu, Haichao Shi, Qiang Liu, Shu Wu, Xingcheng Xu and Xiao-Yu Zhang 15050

Telling Incomplete Truths: Measuring Omission-Induced Misleadingness in Aligned Language Models
Minje Kim, Keeheon Lee and Chi-Guhn Lee 15068

Jais 2: A Family of Arabic-Centric Open Large Language Models
Mohamed Anwar, Abed Alhakim Freihat, George Ibrahim, Mostafa Awad, Abdelrahman Sadallah, Gurpreet Gosal, Gokulakrishnan Ramakrishnan, Sarath Chandran, Biswajit Mishra, Rituraj Joshi, Ahmed Frikha, Etienne Goffinet, Abhishek Maiti, Ali El Filali, Sarah AlBarri, Samujjwal Ghosh, Rahul Pal, Parvez Mullah, Awantika Shukla, Sajid Siddiki, Samta Kamboj, Onkar Pandit, Sunil Kumar Sahu, AbdelRahman Elbadawy, Amr Mohamed, Ahmad Chamma, Evan Dufraisse, Abdelaziz Bounhar, Dani Bouch, Hadi Abdine, Guokan Shang, Fajri Koto, Yuxia Wang, Zhuohan Xie, Ali Mekky, Rania Elbadry, Sarfraz Ahmad, Momina Ahsan, Omar El Herraoui, Daniil Orel, Hasan Iqbal, Kareem Elzeky, Mervat Abassy, Kareem Elozeiri, Saadeldine Eletter, Farah Atif, Nurdaulet Mukhituly, Haonan Li, Xudong Han, Aaryamonvikram Singh, Zainul Abedien Ahmed Quraishi, Neha Sengupta, Larry Murray, Avraham Sheinin, Joel Hestness, Natalia Vassilieva, Hector Xuguang Ren, Zhengzhong Liu, Michalis Vazirgiannis and Preslav Nakov 15090

WeaveBench: A Long-Horizon, Real-World Benchmark for Computer-Use Agents with Hybrid Interfaces
Wanli Li, Bowen Zhou, Yunyao Yu, Zhou Xu, Yifan Yang, Dongsheng Li and Caihua Shan15152

AgentJudgeBench: A Multi-Difficulty Benchmark for Evaluating LLM Judges on Agentic Tool-Calling
Abhigya Verma, Amit Kumar Saha, Seganrasan Subramanian and Sai Harshitha Aluru . . . 15171

MelonBench: A Benchmark for Coded Entity Resolution over Noisy Web Evidence
Zhishan Lin, Jingxiao Li, Yanbing Zhang, Ziyan Liu, Yuqing Shao, Yunfei Tong, Ganghao Liu, Qian Zhang and Zhe Wang 15202

All Relations at Once: Efficient Open-World Tabular Relation Discovery with Hierarchy-Aware Reward Shaping
Xinyu Zhang, Aibo Song, Jingyi Qiu and Tianbo Zhang 15234

Rubric Spans are Label Representations: Joint LLM Encoding for Short Answer Scoring
Zhifan Sun, Sebastian Gombert, Fabian Zehner, Leon Camus, Longwei Cong and Hendrik Drachsler 15246

Beyond Static Symbols: Evidence-Grounded Visual Checklists for Cultural Activity Generation
Hongcan Deng, Bo Ma, Yating Yang, Rong Ma, Zhen Wang, Xi Zhou, Rui Dong and Turghun Osman 15267

The RAT: A Unified Bayesian Model for RAG Evaluation
Pius von Däniken, Felix Matthias Saaro, Mark Cieliebak and Jan Milan Deriu 15295

NaijaS2ST: A Multi-Accent Benchmark for Speech-to-Speech Translation in Low-Resource Nigerian Languages
Marie Maltais, Yejin Jeon, Min Ma, Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Maryam Ibrahim Mukhtar, Daud Abolade, Okepefi Joel Monday, Johnson Sewedo Okepefi and David Ifeoluwa Adelani 15311

Benchmarking and Boosting Multilingual Capabilities of LVLMs via OCR-Centric Reinforcement Learning
Junyuan Gao, Jiahe Song, Jiang Wu, Runchuan Zhu, Shenguanlin, Shasha Wang, Xingjian Wei, Haote Yang, Weijia Li, Bin Wang, Lijun Wu and Conghui He 15335

Alice: A Large-Scale German Benchmark for Rubric-Based Multi-Dimensional Automatic Short Answer Scoring
Zhifan Sun, Sebastian Gombert, Jannik Lossjew, Tobias Wyrwich, Berrit Katharina Czinczel, David Bednorz, Marcus Kubsch, Knut Neumann and Hendrik Drachsler 15366

A Wolf in Sheep's Clothing: Targeted Routing Hijacking in Federated RAG
Junjie Mu and Qiongxiu Li 15388

Do Large Language Models Adapt Politeness Like Humans?
Tianfang Chen, Yuyang Yao and Liping Tang 15412

ReasoningFlow: Discourse Structures for Understanding LLM Reasoning Traces
Jinu Lee, Shivam Agarwal, Amruta Parulekar, Siddarth Madala, Dilek Hakkani-Tür and Julia Hockenmaier 15430

Unlocking Fine-Grained Translation Quality Estimation in LRMs through Mutually Boosting Implicit and Explicit Reasoning
Renfei Dang, Xinye Wang, Zhejian Lai, Weilu Xu, Shimin Tao, Daimeng Wei, Min Zhang and Shujian Huang 15464

MDForge: Agentic Molecular Dynamics Pipeline Design under Sparse Simulator Feedback
Zehong Wang, Yijun Ma, Connor R. Schmidt, Tianyi Ma, Weixiang Sun, Ziming Li, Xiaoguang Guo, Chuxu Zhang, Matthew J. Webber and Yanfang Ye 15482

Encoder-only and decoder-only models exhibit distinct patterns of linguistic bias
Suet-Ying Lam, Qingcheng Zeng and Rob Voigt 15508

Source-Grounded Semantic Reinforcement Learning for Low-Resource Target-Language Generation
Zeli Su, Ziyin Zhang, Pan Zewei, Zhou Liu, Dingcheng Huang, Dehan Li, Zhankai Xu, Longfei Zheng, Xiaolu Zhang, Jun Zhou and Wentao Zhang 15521

CWoMP: Morpheme Representation Learning for Interlinear Glossing
Morris Alper, Enora Rice, Bhargav Shandilya, Alexis Palmer and Lori Levin 15535

Gavel: Agent Meets Checklist for Evaluating LLMs on Long-Context Legal Summarization
Yao Dou, Benjamin Mamut and Wei Xu 15553

The Past Is Prologue: A Plug-in Controller for Selective Updates in Sequentially Evolving LLM Memory
Zihan Chen, Songwei Dong, Chengshuai Shi, Peng Wang, Song Wang, Cong Shen and Jundong Li 15604

ReFeedGraph: Graph-Grounded Reusable Feedback for Debate Improvement
Priya Pitre, Omkar J. Joshi, Timon Ziegenbein and Henning Wachsmuth 15621

Evaluating the Hidden Costs of Personalization in Large Language Models
Yumeng Wang, Yuchen Wu, Cheng Qian, Zhiyuan Fan, Hyeonjeong Ha, Shujin Wu, Jiayu Liu, Heng Ji and Ge Wang 15641

Beyond Output Correctness: Benchmarking and Evaluating Large Language Model Reasoning in Coding Tasks
Yuangang Li, Justin Tian Jin Chen, Ethan Yu, David Hong and Iftekhar Ahmed 15666

A Large-Scale Multi-Dimensional Empirical Study of LLMs for Conversation Summarization
Weixiao Zhou, Gengyao Li, Junnan Zhu, Xianfu Cheng, Feifei Zhai and Zhoujun Li 15690

Geometry First: Shaping Sentence Embeddings with Controlled Singular Value Regularization
Senyu Li, Hao Yu, Ziyuan Wang and David Ifeoluwa Adelani 15711

Select the Key, Then Generate the Rest: Improving Multi-Modal Learning with Limited Data Budget
Xinyu Zhao, Kaishen Wang, Jiaye Yang, Jie Peng, Lirong Xiang, Faizan Siddiqui, Kwonjoon Lee, Behzad Dariush and Tianlong Chen 15727

GUI-360° : A Comprehensive Dataset and Benchmark for Desktop Productivity Computer-Using Agents
Jian Mu, Chaoyun Zhang, Chiming Ni, Lu Wang, Bo Qiao, Kartik Mathur, Qianhui Wu, Yuhang Xie, Xiaojun Ma, Mengyu Zhou, Si Qin, Liqun Li, Yu Kang, Minghua Ma, Qingwei Lin, Saravan Rajmohan and Dongmei Zhang 15742

ObGynLongBench: Revealing the Evidence-to-EHR Gap in Longitudinal EHR Decision-Making
Jun Xiang, Zhijie Bao, Hu Rong, Kaizhou Qin, Wei Chen and Zhongyu Wei 15764

Quantifying Hidden Salt for Precision Healthcare: Sodium Assessment via Joint-Factor Retrieval and Chain-of-Thought Inference
Mingyu Huang, Weiqing Min, Yuehui Fang, Yuna He and Shuqiang Jiang 15789

QUACK: Questioning, Understanding, and Auditing Communicated Knowledge in Multimodal Social Deduction Agents
Ye Yuan, Rui Song, Weien Li, Zeyu Li, Haochen Liu, Xiangyu Kong, Changjiang Han, Yonghan Yang, Zichen Zhao, Zixuan Dong, Fuyuan Lyu, Bowei He, Haolun Wu, Jikun Kang and Xue Liu15805

Ideation Arena: Evaluating LLM-driven Research Ideation with Battle-style Human Expert Annotation
Zhiyu Chen, Keyu Zhao, Jigao Fu, Dong Liang, Yanbiao Wu, Jiaoyang Li, Haidong Xue, Xinhua Zeng, Yuanyi Zhen, Fengli Xu and Yong Li 15825

Bridging Generation and Parsing: Collaborative Self-Distillation for Zero-shot Multi-Intent Spoken Language Understanding
Aziguli Wulamu, Ao He, Dezheng Zhang, Jun Li, Xinyuan Qian, Zhihong Zhu, Libo Qin and Bowen Xing 15850

Some Tokens Behave like Magnets: Revealing Linguistic Organization in the Layers of Language Models
Andrew Liu, Devan Srinivasan and Gerald Penn 15867

Target-Binding Hallucination: A Position-Controlled Diagnostic for Multimodal LLMs in Multi-Image Contexts
Qichao Ma, Yongtai Liu and Xuan Zhu 15887

TukaBench: A Culturally Grounded Jailbreak Benchmark for African Languages
Victor Olalekan Akinode, Senyu Li, Wassim Hamidouche, Syed Waqas Zamir, Inbal Becker-Reshef and David Ifeoluwa Adelani 15907

LogicGraph: Benchmarking Multi-Path Logical Reasoning via Neuro-Symbolic Generation and Verification
Yanrui Wu, Lingling Zhang, Xinyu Zhang, Jiayu Chang, Pengyu Li, Xu Jiang, Jingtao Hu and Jun Liu 15931

TEMPO: Temporally-grounded Multi-task Post-training for Large Audio-Language Models
Apoorva Kulkarni, Kaousheik Jayakumar, Sreyan Ghosh, Utathya Aich, Ramani Duraiswami and Dinesh Manocha 15960

Text as a Universal Interface for Transferable Personalization
Yuting Liu, Jian Guan, Jia-Nan Li, Wei Wu, Jiang-Ming Yang, Jianzhe Zhao and Guibing Guo 15977

Evaluating LLM-Simulated Conversations in Modeling Inconsistent and Uncollaborative Behaviors in Human Social Interaction
Ryo Kamoi, Ameya Godbole, Binglin Zhou, Xiaoxin Lu, Longqi Yang, Rui Zhang, Mengting Wan and Pei Zhou 16019

PhoenixNest-Video: Evidence-Grounded Multimodal Agent Framework for Automated Video Interview Assessment
Yuxuan Fan, Miaojun Huang, Haimei Zhang, Jingshen Wu and Hao Liu 16045

Internalizing the Future: A Unified Agentic Training Paradigm for World Model Planning
Xuan Zhang, Zhijian Zhou, Lingfeng Qiao, Yulei Qin, Ke Li, Xing Sun, Xiaoyu Tan, Chao Qu and Yuan Qi 16068

Structural Uncertainty from Self-Preference in Black-Box Language Models
Baishali Chaudhury, Mengdie Flora Wang, Hyunji Hayley Park, Rahul Ghosh, Sungmin Hong and Jae Oh Woo 16119

Would you still call this Dax? Novel Visual References in VLMs and Humans
Ada Defne Tur, Gaurav Kamath, Joyce Chai, Siva Reddy and Benno Krojer 16148

The Dialect Tax: Dialectal Biases Persist throughout the Language Modeling Pipeline
Elle 16184

SAGA: Structured Applicability-Guided Alignment for Conversational Legal Retrieval
Xinxuan Lv, Minghan Li and Guodong Zhou 16207

SkillSafetyBench: Evaluating Agent Safety under Skill-Facing Attack Surfaces
Chang Jin, An Wang, Zeming Wei, Kai Wang, BiaoJie Zeng, Qiaosheng Zhang, Chao Yang, Jingjing Qu, Xia Hu and Xingcheng Xu 16227

Implicit vs. Explicit Prompting Strategies for LVLMs in Referential Communication
Peter Zeng, Amie J. Paige, Weiling Li, Susan Brennan, Owen Rambow and Cameron Robert Jones 16269

Lifelong Intent Discovery: A Continual Learning Protocol and Method
Peng Chen, Yu Xia and Bang Wang 16286

HugAgent: A Human Simulation Benchmark for Individual-Level Reasoning
Chance Jiajie Li, Zhenze Mo, Yuhan Tang, Ao Qu, Jiayi Wu, Kaiya Ivy Zhao, Yulu Gan, Jie Fan, Jiangbo Yu, Hang Jiang, Paul Pu Liang, Jinhua Zhao, Luis Alberto Alonso Pastor and Kent Larson 16315

Stateful Visual Encoders for Vision-Language Models
Zirui Wang, Junwei Yu, Adam Yala, David M. Chan, Joseph E. Gonzalez and Trevor Darrell 16377

Measuring, Localizing, and Reducing the AI-Like Writing Style of Post-Trained LLMs
Aniket Anand, Janvijay Singh, Zhewei Sun, Dilek Hakkani-Tür and Nick Feamster 16400

Agatha: Gated Hopfield Attention for Adversarially Robust Natural Language Processing
Andreia Podasca and Anup Das 16437

VIB-Simp: IB-Inspired Output-Logit Regularization for German Text Simplification
Mofan Duan, Yanyan Yang and Xiufeng Liu 16483

SHERLOC: Structured Diagnostic Localization for Code Repair Agents
Hovhannes Tamoyan, Sean Narenthiran, Erik Arakelyan, Mira Mezini and Boris Ginsburg 16496

MisconceptionBench: Evaluating LLMs on Science Misconceptions and Student Response Distributions
Ying Zhang, Zhaoji Wang, Chenyu Zhang, Yizhu Gao, Wenchao Ma and Xiaoming Zhai . 16525

From Consistency to Complementarity: Aligned and Disentangled Multi-modal Learning for Time Series Understanding and Reasoning
Hang Ni, Weijia Zhang, Fei Wang, Zezhi Shao and Hao Liu 16540

Diagnosing LLM Arbitration Behavior over Pre-evidence Epistemic States in RAG-based Fact-Checking
Yuxi Sun, Wenbo Shang, Wei Gao, Xin Huang and Jing Ma 16563

Align Then Adapt: Label-Efficient Adapter Learning for Asymmetric Dense Retrieval
Seiji Maekawa, Moin Aminnaseri, Pouya Pezeshkpour and Estevam Hruschka 16595

M^2Note: Continual Evolution of Vision Language Models via Mistake Notebook Learning
Haiwen Li, Jing Tang, Rui Chen, Lei Sun, Zining Chen and Xiangxiang Chu 16611

Wrong Prediction, Right Answer: Recovering Evidence from Collapsed LLM Sequence Scores
Qiyao Yan, Chenpeng Wang and Liangming Pan 16630

Extended Inductive Reasoning for Personalized Preference Inference from Behavioral Signals
Jia-Nan Li, Jian Guan, Wei Wu and Rui Yan 16650

MIRA: A Bilingual Benchmark for Medical Information Response Audit
Mengyu Xu, Qiaoxin Yang, Qianqian Wang, Xiwei Dai, Weiyi Wu and Chongyang Gao . . 16683

Lexical Perturbations Disrupt LLM Reasoning: An Empirical Study of Attention Diversion
Jiaqian Zhu, Yang Zhang, Junhua Ding and Xiaowei Yu 16706

HELM: Steering Long-Horizon Agents with Learned Hierarchical Memory and Epistemic Governance
Jialong Liu, Yuling Shi, Zuchao Li and Ping Wang 16728

Why It Hurts: Identifying the Drivers of Negative Thoughts in Emotional Support Conversations
Hainiu Xu, Zhaoyue Sun, Hanqi Yan, Jinhua Du, Caroline Catmur and Yulan He 16754

TRACES: Proactive Safety Auditing for Multi-Turn LLM Agents via Trajectory-State Modeling
Jiaqian Li, Yanshu Li, Boxuan Zhang, Ruixiang Tang and Kuan-Hao Huang 16798

Safety Sentry: Context-Aware Human Intervention via EXECUTE–ASK–REFUSE Routing
Tianyu Chen, Chujia Hu and Wenjie Wang 16826

Citing Less Critically: LLMs Reshape the Rhetoric and Reach of Scientific Citation
Yixuan Liu, Lin Chen, Zhuoqi Liu, Jianglin Lu and Dakota Murray 16853

SIMPA: Restricting What Experts See, Not What They Produce, in MoE-PEFT
Pham Tri Quang, Khoa Tran, Giang Do, Hung Le, Truyen Tran and Tho Quan 16873

INFA-Guard: Mitigating Malicious Propagation via Infection-Aware Safeguarding in LLM-Based Multi-Agent Systems
Yijin Zhou, Xiaoya Lu, Dongrui Liu, Junchi Yan and Jing Shao 16895

QUBRIC: Co-Designing Queries and Rubrics for RL Beyond Verifiable Rewards
Rongzhi Zhang, Rui Feng, Zhihan Zhang, Jingfeng Yang, Qingyu Yin, Xin Liu, Zixuan Zhang, Priyanka Nigam, Bing Yin, Tuo Zhao and Chao Zhang 16914

Learning to Evict: Policy-Driven KV Cache Optimization for Multi-Hop Multimodal Spatial Reasoning
Jiale Chang and Ying Li 16932

Mitigating Forgetting Between Supervised and Reinforcement Learning Yields Stronger Reasoners
Xiangchi Yuan, Xiang Chen, Tong Yu, Dachuan Shi, Can Jin, Wenke Lee and Saayan Mitra 16948

From Utterances to Networks: Modelling Slang Adoption and Diffusion Across Subreddits
Xiaoning Wang, Ted Underwood and Zhewei Sun 16972

The Emergence and Sparsification of a Syntactic Circuit
Qing Yao, Sasha Boguraev, Tiago Pimentel and Kyle Mahowald 16987

Rethinking LoRA Memory Through the Lens of KV Cache Compression
Chunsheng Zuo, Liaoyaqi Wang, William Jurayj, William Fleshman and Benjamin Van Durme 17013

Half a Link can Be Enough to Predict a Whole Link: Understanding Generalization in Knowledge Graph Foundation Models
Cosimo Gregucci, Obaidah Theeb, Daniel Hernández, Antonio Vergari and Steffen Staab . 17031

Rating the Raters: Rasch Measurement Theory for LLM Evaluation
Pratik S. Sachdeva and Nathan Boudol 17056

EmoTra-TTS: Smooth Intra-Utterance Emotion Transitions for Speech Synthesis
Tianchi Liu, Zeyang Song, Tianrui Wang, Zhipeng Li, Chenglin Xu and Steven Y. Guo . . . 17073

FlexLTS: A Unified Paradigm for Content-Accurate Lip-to-Speech Synthesis with Optional Lip Reading
Zhechen Liu, Shuang Yang, Shiguang Shan and Xilin Chen 17092

Handoff Debt: The Rediscovery Cost When Coding Agents Take Over Interrupted Tasks
Dipesh KC and Anjila Budathoki 17110

SurgMind-R1: Overcoming Reasoning Collapse in Surgical Vision-Language Models via Reinforcement Learning
Yifan Luo, Guoming Wang, Ling Xiao, Juncheng Li and Siliang Tang 17127

BreastGPT: A Multimodal Large Language Model for the Full Spectrum of Breast Cancer Clinical Routine
Yang Liu, Jiajin Zhang, Danyang Tu, Yaojun Hu, Jiao Qu, Jiuyu Zhang, Yu Shi, Wei Fang, Shi Gu, Ling Zhang and Yingda Xia 17144

XCR-Bench: Benchmarking Cross-Cultural Reasoning in LLMs via Culture-Specific Items and Hall's Triad
Mohsinul Kabir, Tasnim Ahmed, Md Mezbaur Rahman, Shaoxiong Ji, Hassan Alhuzali, Yuechen Jiang, Jimin Huang and Sophia Ananiadou 17179

Portuguese-IFEval: Instruction-Following Evaluation in Portuguese
Emanuel Borges Passinato, Gustavo Pires Fontana, Cibelle Pires Botelho, Júlia Jaime Biancolini, Diogo Fernandes Costa Silva and Arlindo Rodrigues Galvão Filho 17210

Beyond Truncation: Rethinking LLM Decoding as Ensemble Pruning
Dunyao Xue, Chengshuo Du, Zhengbo Wang, Wenlin Dai and Cheng Meng 17229

DrugReason: Dynamic Multi-View Reasoning over Knowledge Graph and Language Evidence for Drug Repurposing
Zijie Liu, Hongxuan Li, Zhen Tan, Jinhao Duan, Baixiang Huang, Zunpeng Liu, Kai Shu and Tianlong Chen 17254

Beyond Magnitude: Contrastive Routing for Modular Mixture-of-Experts
Nikolaos Xiros, Dimitrios Damianos, Maria-Eleni Zoumpoulidi, Leon Voukoutis, Vassilis Katsouros and Georgios Paraskevopoulos 17267

LifeAgentBench: Benchmarking LLMs for Long-Horizon, Cross-Dimensional Lifestyle Health Reasoning
Ye Tian, Zihao Wang, Onat Gungor, Xiaoran Fan and Tajana Rosing 17281

Beyond Accuracy: Community Perspectives on Machine Translation
Yujun Wang, Ehud Reiter, Shimei Pan, Steffen Eger and Wei Zhao 17302

An Embarrassingly Simple Detector for Model Extraction Attacks in Large Language Model API Traffic
Shuze Liu, Qianwen Guo and Yushun Dong 17320

Accelerating Constrained Decoding with Token Space Compression
Michael Sullivan and Alexander Koller 17339

From National Curricula to Cultural Awareness: Constructing Open-Ended Culture-Specific Question Answering Dataset
Haneul Yoo, Won Ik Cho, Geunhye Kim and Ji Yoon Han 17353

Toward Latent Language Model Skills Steering and Optimization: An Empirical Study
Xunyi Jiang, Junda Wu, Yuxin Xiong, Sheldon Yu, Tong Yu, David Arbour, Ritwik Sinha, Julian McAuley and Hongyi Wen 17373

LEAP: Likelihood Elicitation and Aggregation for LLM-based Probabilistic Forecasting
Yufei Chen, Yiran Zhao, Xiaogang Xu, Qipeng Xie, Jiafei Wu and Zhe Liu 17393

Implicit Bias in LLMs for Transgender Populations
Micaela Hirsch, Marina Elichiry, Blas Radi, Tamara Quiroga, David Restrepo, Luciana Benotti, Veronica Xhardez, Jocelyn Dunstan and Enzo Ferrante 17413

On the (In)effectiveness of AMR Augmentation for Large Language Models
Hoa Quynh Nhung Nguyen, Jacopo Staiano and Michael Sullivan 17428

The Cognitive Bandwidth Bottleneck: Shifting Long-Horizon Agents from Planning with Actions to Planning with Schemas
Baixuan Xu, Tianshi Zheng, Zhaowei Wang, Hong Ting Tsang, Weiqi Wang, Tianqing Fang and Yangqiu Song 17451

Masking Stale Observations Helps Search Agents — Until It Doesn't: A Regime Map and Its Mechanism
Haoxiang Zhang, Qixin Xu, Zhuofeng Li, Lei Zhang, Pengcheng Jiang, Yu Zhang and Julian McAuley 17469

AfriScience-MT: Towards Decolonizing Science in Africa through Text Translation
Idris Abdulmumin, Tajuddeen Gwadabe, Shamsuddeen Hassan Muhammad, David Ifeoluwa Adelani, Nomonde Khalo, Ibrahim Said Ahmad, Abiodun Modupe, Anina Mumm, Sibusiso Biyela, Michelle Rabie, Johanna Havemann, Marek Rei, Jade Abbott and Vukosi Marivate 17506

Trace2Skill: Distill Trajectory-Local Lessons into Transferable Agent Skills
Jingwei Ni, Yihao Liu, Xinpeng Liu, Yutao Sun, Mengyu Zhou, Pengyu Cheng, Dexin Wang, Erchao Zhao, Xiaoxi Jiang and Guanjun Jiang 17556

Are LLM Decisions Faithful to Verbal Confidence?
Jiawei Wang, Yanfei Zhou, Siddartha Devic and Deqing Fu 17578

Seeing What Is Actually There: PriVE-Bench and PriVE-Tools for Counterfactual Evaluation of Agentic Visual Evidence in VLMs
Jingyu Sun, Jiachen Tu, Yuyang Xue, Yaoxin Jiang, Guoyi Xu, Zhengtao Yao, Rui Qian, Yizheng Sun, Hongpeng Zhou, Jingyuan Sun and Yan Lin 17595

Sample Where You Struggle: Sharpening Base Model Reasoning via Entropy-Guided Power Sampling
Hong Guo, Nianhui Guo, Christoph Meinel and Haojin Yang 17623

Building Auditable AI Text Detectors with Sparse Autoencoders
Liam Dugan, Callum Stuart McDougall, Matthew Denton, Philip Pham, Christine Kaeser-Chen and Chris Callison-Burch 17640

WildSEEK: Evaluating Language Models for Information-Seeking
Tanise Ceron, Joachim Baumann, Elisa Bassignana, Berat Cabuk, Dirk Hovy and Debora Nozza 17690

OLIVIA: Online Learning via Inference-time Action Adaptation
Sheldon Yu, Junda Wu, Xintong Li, Nikki Lijing Kuang, Sizhe Zhou, Tong Yu, Jiawei Han, Jingbo Shang and Julian McAuley 17719

Real-Time Deadlines Reveal Fragile Temporal Adaptation in LLM Strategic Dialogues
Neil Sehgal, Sharath Chandra Guntuku and Lyle Ungar 17735

From Disclosure to Evaluability: an NLP-Based Analysis of Corporate Transition Plans
Tom Calamai, Oana Balalau, Theo Le Guenedal and Fabian M. Suchanek 17757

SONIC-O1: A Real-World Benchmark for Evaluating Multimodal Large Language Models on Audio-Video Understanding
Ahmed Y. Radwan, Christos Emmanouilidis, Hina Tabassum, Deval Pandya and Shaina Raza 17786

Are Benchmarks a Source of Bias in LLMs? MMLU, CommonSenseQA and MGSM under the Microscope
Fanny Ducel, Shayan Ahmed Shariff, Ibrahim Al Kotob, Binesh Arakkal Remesh, Lucie Digoin-Caparros, Aurélie Névéol and Karën Fort 17826

Diagnostic Analysis for Large Language Model Reasoning
Xinyu Liu, Dongqi Fu and Jun Wu 17842

DigitalCoach: Communication and Grounding Gaps in Human and Agentic Computer Use Coaching
Meng Chen, Anya Ji, Tsung-Han Wu, Tobias Maringgele, David M. Chan, Alane Suhr and Amy Pavel 17868

From Saliency Ranking to Evidence Preservation: Visual Token Condensation for Efficient Multimodal Reasoning
Xin Zou, Di Lu, Yibo Yan, Shuliang Liu, Yizhou Wang, Xiang Liu, Linfeng Zhang and Xuming Hu 17908

UserBench: An Interactive Gym Environment for User-Centric Agents
Cheng Qian, Zuxin Liu, Akshara Prabhakar, Zhiwei Liu, Jianguo Zhang, Haolin Chen, Heng Ji, Weiran Yao, Shelby Heinecke, Silvio Savarese, Caiming Xiong and Huan Wang 17936

TROPE: Typed Representations and Operators for Problem Editing in LLM Reasoning
Vadim Bikbulatov and Ovanes Petrosian 17967

DYNAMO: Dynamic Skill-Tool Evolution for Vision-Language Agents
Yutao Sun, Yanting Miao, Hao-Xuan Ma, Mengyu Zhou, Mingshuai Chen, Tiancheng Zhao, Dexin Wang, Lei Lv, Li Xu, Xiaoxi Jiang and Guanjun Jiang 17994

Auditing LLM Benchmarks with Item Response Theory
Sander Land and Daniel M. Bikel 18016

Scale or Reason? A Compute-Equivalent Analysis of Reasoning Distillation
Nicolas Boizard, Hippolyte Gisserot-Boukhlef, Kevin El Haddad, Celine Hudelot and Pierre Colombo 18040

PyraVid: Hierarchical Multimodal Memory for Long-Horizon Video Reasoning
Sikuan Yan, Sicheng Dong, Haotong Wang, Ercong Nie, Yilun Liu, Jinhe Bi, Yingjie Xu, Susanna Schwarzmann, Riccardo Trivisonno, Volker Tresp and Yunpu Ma 18065

EST-PRM: Stress-Testing Process Reward Models Before They Become Load-Bearing
Ibne Farabi Shihab, Fariya Afrin, Sanjeda Akter and Anuj Sharma 18085

Validity-Aware Jailbreak Evaluation for Large Language Models
Qilong Wu, Sahil Wadhwa, Pranab Mohanty, Giri Iyengar and Varun Chandrasekaran 18109

UserRL: A Gym-Based Testbed for User-Centric Reinforcement Learning
Cheng Qian, Zuxin Liu, Akshara Prabhakar, Jielin Qiu, Zhiwei Liu, Haolin Chen, Shirley Kokane, Heng Ji, Weiran Yao, Silvio Savarese, Caiming Xiong and Huan Wang 18133

FOXGLOVE: Understanding Goal-Oriented and Anchored Writing Feedback from Experts and LLMs on Argumentative Essays
Yijun Liu, Yifan Song, John R. Gallagher, Sarah Sterman and Tal August 18164

CALIBURN: Self-Calibrated LLM Unlearning Alignment
Zhengbang Yang, Yisheng Zhong, Junyuan Hong and Zhuangdi Zhu 18186

Camellia: Benchmarking Cultural Biases in LLMs for Asian Languages
Tarek Naous, Anagha Savit, Carlos Rafael Catalan, Geyang Guo, Jaehyeok Lee, Kyungdon Lee, Lheane Marie Dizon, Mengyu Ye, Neel Kothari, Sahajpreet Singh, Sarah Masud, Tanish Patwa, Trung Thanh Tran, Zohaib Khan, Alan Ritter, Tanmoy Chakraborty, Yuki Arase, Keisuke Sakaguchi, JinYeong Bak and Wei Xu 18203

PhysWave: Physics-Guided Latent Diffusion Models for Controllable Spatial Audio Generation
Lingfeng Yao, Chenpei Huang, Xingke Yang, Ziye Geng, Changqing Luo, Hao Wang, Jiang Liu and Miao Pan 18227

Spectral DPO: Mitigating Validation Collapse via Heavy-Tailed Regularization
Ibne Farabi Shihab, Sanjeda Akter and Anuj Sharma 18243

Conformalized Large Language Models under Configuration Shift
Yuqicheng Zhu, Jialin Yu, Lin Li, Gengyuan Zhang, Zhen Yang, Steffen Staab, Puneet K. Dokania, Philip Torr, Jie Tang and Evgeny Kharlamov 18261

PreUnlearn: Auditing Collateral Knowledge Damage Before Large Language Model Unlearning
Bo Su, Ankit Shah and Thai Le 18290

BALMS: Benchmarking Agentic LLMs for Longitudinal Mental Health Sensing
Yu Yvonne Wu, Arvind Pillai, Yuliang Chen, Yuwei Zhang, Sudarshan Regmi, Tess Z Griffin, Michael V. Heinz, Lisa Marsch, Nicholas C. Jacobson and Andrew Campbell 18302

More Haste, Less Speed: Weaker Single-Layer Watermark Improves Distortion-Free Watermark Ensembles
Ruibo Chen, Yihan Wu, Xuehao Cui, Jingqi Zhang and Heng Huang 18322

One Frozen Simulator Is Not Enough: Simulator Collapse in Multi-Agent RL
Simon Yu, Nicholas Tomlin, Marwa Abdulhai, Derek Chong, Ximing Lu, Abe Bohan Hou, Dilara Soylu, Sergey Levine, Christopher D Manning and Weiyan Shi 18336

ChronoVision: Temporal Reasoning via Latent State Reconstruction
Yifan Shen, Jian Xu, Boyi Li, Yuner Zhang, Tianjiao Yu, Bingxuan Li, Houze Yang, Rushi Wang and Xu Cao 18374

LexRubric: A Rubric-Guided Diagnostic Benchmark for Open-Ended Legal Tasks
Yifan Chen, Haitao Li, Yiran HU, Kaisong Song, Jun Lin, Yueyue Wu, Qingyao Ai, Min Zhang and Yiqun Liu 18406

TaskPress: Query-Agnostic KV Cache Compression via Task-Guided Pruning
Wonpyo Park and Seung-Won Hwang 18423

Labels have Human Values: Value Calibration of Subjective Tasks
Mohammed Fayiz Parappan and Ricardo Henao 18435

When Evolution Meets Momentum: Orchestrating Goal-oriented and Process-oriented reasoning for LLM Inference Scaling
Mingyu Jin, Haifeng Chen, Yuanzhou Chen, Jonathan Light, Wenchao Yu and Wei Cheng 18454

AdaMame: A Training Recipe for Adaptive Multilingual Reasoning
Dayeon Ki, Kevin Duh and Marine Carpuat 18490

SCRIBE: Learning Reusable Skill Rubrics for Process Reward Models
Yuxuan Jiang and Francis Ferraro 18512

Rethinking the Evaluation of Efficiency Methods for Multi-Agent Systems
Jiamu Zhang, Lingxi Zhang, Pengjun Lu, Qiyue Zhang, Yu-Neng Chuang, Zhengchen Li, Shuai Xu, Vipin Chaudhary and Hanjie Chen 18527

ALERT: Zero-shot LLM Jailbreak Detection via Internal Discrepancy Amplification
Xiao Lin, Philip Li, Zhichen Zeng, Ting-Wei Li, Tianxin Wei, Xuying Ning, Gaotang Li, Yuzhong Chen and Hanghang Tong 18563

Can LLMs Take the Pulse of the Economy? A Real-Time Evaluation of LLM Nowcasts on Macroeconomic Indicators
Xinyue Zhao, Ruiyi Zhang, Liqin Ye, Rui Cao, Pengtao Xie and Sudheer Chava 18583

Monkey Jump: MoE-Style PEFT for Efficient Multi-Task Learning
Nusrat Jahan Prottasha, Md Kowsher, Chun-Nam Yu, Chen Chen and Ozlem Garibay 18631

MM-COMPLIANCE: A Multilingual Multimodal Dataset for Sustainability Compliance Auditing
Rui He, Jianhua Yang, Tomoki Kera, Atsushi Keyaki, Anaïs Lhuissier, Juyeon Kang, Min-Yuh Day, Guan-Bo Yang, Hsin-Hsi Chen, Dittaya Wanvarie, Yongjae Lee, Suhwan Park, Yuriko Nakao, Chung-Chi Chen and Yohei Seki 18675

Integrated and Cross-Architecture Interpretation of LLM Reasoning
Leonardo Matthew Yauw, Wei-Bin Kou and Yujiu Yang 18691

MemUse: Moving Memory Evaluation from Direct QA to Natural Integration in Long-Term Human-AI Conversation
Ryuichi Sumida, Koji Inoue and Tatsuya Kawahara 18719

TimePrune: Budgeted Temporal Pruning for Efficient Multimodal Time Series Analysis
Chenghao Liu, Qian Liu, Aniket Mahanti, Jindong Han, Min Li and Wei Fan 18750

Do Language Models Understand Implicit Logical Meaning? A Case Study of Scope Ambiguity Resolution in Context
Jieun Kim, Suhyeon Park, Kunhui Lee and Seung-Hoon Na 18772

StratPrune: Cascade Stratum Pruning for Efficient Time Series Reasoning
Chenghao Liu, Qian Liu, Aniket Mahanti, Alex Xing Wang, Jingwei Yi and Wei Fan 18796

HumorGen: Cognitive Synergy for Humor Generation in Large Language Models via Persona-Based Distillation
Edward Ajayi and Prasenjit Mitra 18814

TOUCAN: Synthesizing 1.5M Tool-Agentic Data from Real-World MCP Environments
Zhangchen Xu, Adriana Meza Soria, Shawn Tan, Anurag Roy, Ashish Sunil Agrawal, Radha Poovendran and Rameswar Panda 18842

Batch-of-Thought: Cross-Instance Learning for Enhanced LLM Reasoning
Xuan Yang, Furong Jia, Roy Xie, Xi Xiong, Hengwei Bian, Jian LI and Monica Agrawal . . 18873

KnowMeBenchV2: Evidence-Grounded Person-Centric Long-Video Understanding
Tingyu Wu, Guangyu Cao, Qizhen Lan, Zhisheng Chen, Ziyan Weng, Bingkun Zhu, Miao Su, Chenglong Li, Zhengwei Xie, Huacan Wang, Sen Hu, Haocun Ye, Nan AN, Zaoqu Liu and Ronghao Chen 18893

Secret Stealing Attacks on Local LLM Fine-Tuning through Supply-Chain Model Code Backdoors
Zi Li, Tian Zhou, Wenze Li, Jingyu Hua, Yunlong Mao and Sheng Zhong 18914

EGR-FV: Evidence-Grounded Remixing for Debiased Fact Verification via Counterfactual Necessity-Guided Training
Jun Yang, Dandan Song, Zhijing Wu, Yuhang Tian, Huipeng Ma, Chenhao Li, Yizhou Jin, Guangyuan Feng and Xudong Li 18931

SPEC-Distill: Towards Efficient Speculative Distillation for Reinforcement Learning with Verifiable Rewards
Bingshuai Liu, Zijun Min, Ante Wang, Jiajun Cao, Haibo Zhang, Liang Yao, Shuman Liu, Anxiang Zeng and Jinsong Su 18946

Can Generalist Agents Automate Data Curation?
Feiyang Kang, Hanze Li, Adam Nguyen, Mahavir Dabas, Jiaqi W. Ma, Frederic Sala, Dawn Song and Ruoxi Jia 18960

Attention Amnesia in Hybrid LLMs: When CoT Fine-Tuning Breaks Long-Range Recall, and How to Fix It
Xinyu Zhou, Boyu Zhu, Yi Xu, Zhiwei Li, Yingfa Chen, Huiming Wang and Zhijiang Guo 18996

Translation-Induced Label Drift in Natural Language Inference: Evidence from Nine Languages
Muhammad S. Abdo, Phakphum Artkaew, Sadia Ashraf, Elder Henrique Attala E Paiva, Jose Benavides, Hale Karkoutli, Melissa Lieffers, Elias Naske, Wenchen Shi, Jinyuan Wang and Sandra Kübler 19014

ConsensusBench: Benchmark of Consensus Nodes for LLM Reasoning via Outcome Reward Densifying
Shi-Qi Yan, Chao-Hong Tan, Qian Chen, Wen Wang, Xiangang Li and Zhen-Hua Ling . . . 19028

The Art of Efficient Reasoning: Data, Reward, and Optimization
Taiqiang Wu, Zenan Xu, Bo Zhou and Ngai Wong 19042

VectorGym: A Multi-Task Benchmark for SVG Code Generation, Sketching and Editing
Joan Rodriguez, Haotian Zhang, Abhay Puri, Haoran Dai, Tianyang Zhang, Meng Lin, Rishav Pramanik, Xiaoqing Xie, Marco Terral Rodriguez, Darsh Kaushik, Aly Shariff, Sai Rajeswar, Perouz Taslakian, Spandana Gella, David Vazquez, Christopher Pal and Marco Pedersoli 19064

Skill Reuse as Compression in Agentic RL
Zhikun Xu, Yu Feng, Jacob Dineen, Taiwei Shi, Jieyu Zhao and Ben Zhou 19092

PunGraph: Retrieval-Enhanced Phonetic-Semantic Graph Reasoning for Pun Understanding
Yuchen Su, Zijian Huang, Yaotian Shi, Shaoxin Zhong, Ruofan Wang, Mengze Li, Yonghua Zhu, Diana Benavides-Prado and Michael J. Witbrock 19111

Token-weighted Direct Preference Optimization with Attention
Chengyu Huang, Zhuohang Li, Sheng-Yen Chou and Claire Cardie 19125

Dimple: Discrete Diffusion Parallel Generation for Large Multimodal Modal
Runpeng Yu, Xinyin Ma and Xinchao Wang 19146

SUNDER: Selective Unmasking for Text Understanding
Alphaeus Dmonte, Tharindu Ranasinghe and Marcos Zampieri 19171

CAPTR: Caption-based Context Pruning for Multimodal Tabular Reasoning
Prasham Titiya, Leon Engländer, Gaurav Najpande, Tampu Ravi Kumar, Iryna Gurevych and Vivek Gupta 19189

Which Metrics Save the Most Human Annotation? Prediction-Powered Evaluation and Meta-Evaluation
Mingqi Gao, Anthony Sicilia and Weiyan Shi 19215

Beyond Semantic Organization: Memory as Execution State Management for Long-Horizon Agents
Yaoqi Chen, Haibin Lai, Yuru Feng, Chuyu Han, Qianxi Zhang, Baotong Lu, Menghao Li, Xinjiang Wang, Zewen Jin, Hao Wu, Cheng Li and Qi Chen 19255

CoVer: Conflict-aware Claim Verification
Shuning Zhang, Dai Shi, Bohao Chu, Hui Wang, Yuwei Chuai, Yifan Wang, Jingruo Chen, Simin Li, Xin Yi and Hewu Li 19273

From Debate to Decision: Conformal Social Choice for Safe Multi-Agent Deliberation
Mengdie Flora Wang, Haochen Xie, Guanghui Wang, Aijing Gao, Guang Yang, Ziyuan Li, Qucy Wei Qiu, Fangwei Han, Hengzhi Qiu, Yajing Huang, Bing Zhu and Jae Oh Woo 19290

ABench: Adversarily Benchmarking Large Language Models across Taxonomized Values
Liping Yang, Yongqi Leng, Jingting Zheng, Yuqi Ren, Shaolin Zhu and Deyi Xiong 19313

Behind-the-Scenes Challenges in Sino-Tibetan Low-Resource Language Technology Development
Shuheng Liu and Michael L. Best 19329

Domain-Specific Jargon in Large Language Models: A Comparative Analysis between General-Purpose and Specialist Models
Darin Keng and Zhewei Sun 19343

Diagnosing the Fact-Grounding Gap in Multi-Hop Question Answering
Kevin Mo, Nathan Mo and Richard Zhu 19363

Language Diffusion Models are Associative Memories Capable of Retrieving Unseen Data
Bao Pham, Mohammed J Zaki, Luca Ambrogioni, Dmitry Krotov and Matteo Negri 19377

DUPO: Dual-Mode Unified Policy Optimization for Tool-Using Agents
Zexin Hu, Mu Hu, Lijuan Yang, Yisen Wang, Xiaogang Zhu, KangJian Jiao, Yu Fu, Lihong Gu, Zhiyong Wang and Kun Hu 19405

AutoForge: Automated Environment Synthesis for Agentic Reinforcement Learning
Shihao Cai, Runnan Fang, Jialong Wu, Baixuan Li, Xinyu Wang, Yong Jiang, Liangcai Su, Liwen Zhang, Wenbiao Yin, Zhen Zhang, Pengjun Xie, Xiaobin Wang and Fuli Feng 19422

UT-ACA: Uncertainty-Triggered Adaptive Context Allocation for Long-Context Inference
Lang Zhou, Shuxuan Li, Zhuohao Li, Shi Liu, Wei-Shi Zheng and Zhilin Zhao 19434

ALTSTEER: Selective Safety Steering for Moving Beyond Hard Refusals to Constructive Alternatives
Hoejoon Kwon, Byeonggeuk Lim, Kahyeon Kim and YoungBin Kim 19455

Where Do Multilingual Vision-Language Encoders Fail on Low-Resource Languages?
Donghoon Han, SungHyun Moon, Aidyn Zhakatayev, Junghun Cha and SeungJae Lee . . . 19474

Beyond the Retweet: A Large-Scale Dataset of Reader Emotion, Propagation Intent, and Behavior
Yujin Zheng, Jian Liao, Suge Wang, JianXing Zheng, Deyu Li and Qian Chen 19506

VeriDx: Earning the Right to Diagnose with Disease-Centric Verification
Zhong Cao and Shuying Chen 19543

One Forward Beats Two: InnerZoom for Accurate and Efficient GUI Grounding
Chen Liu, Ling Chen, Hanzhang Zhou, Liangyu Chen, Chenglin Cai, Xin Yu, Steven Hoi and Yue Wang 19560

Between Multiple Choice and Open Response: Evaluating LLMs with Option-Free Deterministic Scoring
ZhongKen Hew, Sze Jue Yang and Chee Seng Chan 19582

ReViCo: Unveiling the Limitations of VLMs in Visual Text Understanding via Error Correction
Bojun Zhang, Junhong Liang, Feifei Zhai, Fengxian Ji and Yu Zhou 19606

You Really Didn't Get That? Benchmarking Social Pragmatic Inference for Indirect and Playful Chinese Online Comments
Shiwei Hong, Junjie Ma, Emma Jiren Wang, Ethan Z. Rong, Siying Hu, Haichang Li, Ziying Wang and Zhicong Lu 19623

Why Can't They Remember? Uncovering Representation and Retrieval Bottlenecks in Multi-Turn Acoustic Memory
Yang Xiao, Siyi Wang, Han Yin, Hong Jia, Vidhyasaharan Sethu, Eun-Jung Holden and Ting Dang 19641

Noisy Test-Time Reinforcement Learning for Code LLMs
Xikai Yang, Hieu Trung Nguyen, Dunyuan Xu, Yuzhi Zhao, Jinpeng Li, Wenao Ma and Pheng-Ann Heng 19656

From Evidence to Exact Answers: Copy-Preserving Compression and Conversion-Aware Calibration for Frozen SLM RAG
Dong Ho Kang, HyeonJeong Cha and Daein Weon 19678

Beyond Passive Debate: Intent-Driven Multi-Agent Deliberation for Reasoning under Incomplete Knowledge Graphs
Yongqiang Liu, Qiyao Peng, Binrong Liu, Jiawei Chen, Xuewei Li and Ning Yuan 19695

Architecture-Conditioned Reinforcement Learning Makes Sliding-Window Attention Competitive in Math Reasoning
Kai Liu, Peijie Dong, Xie Xinchen, Jianfei Gao, Qipeng Guo, Xiaowen Chu, Shaoting Zhang and Kai Chen 19713

BOW: Training Language Models to Reason Over Plausible Next Words
Ming Shen, Zhikun Xu, Jacob Dineen, Xiao Ye and Ben Zhou 19732

SAFE: An LLM-as-Verifier Framework for Evidence-Grounded Multi-Hop Reasoning
Daeyong Kwon, Soyoung Yoon and Seung-Won Hwang 19755

ADE: Agentic Data Evolution Framework for Human-Centered Objectives
Yang Yu, Yilin Jiang, Zexuan Fei, Yiming Luo, Xingkai Song, Kaiyi Huang, Aimin Zhou, Xin Lin and Fei Tan 19789

LongGuard: Mechanistic Analysis and Training-Free Mitigation of Long-Context Failure in Safety Guardrails
Ziyang Chen, Xing W and Songlin Hu 19813

KernelQuant: Post-Training Quantization via Representation Kernel Preservation
JuneYoung Park, Youngwook Kwon, Seongwan Kim and Jaeho Lee 19833

Key Path Identification for Resolving Knowledge Conflicts via SAE-based Steering
Wenbo Zhang, Zhongxiang Sun, Zhiguang Han and Jun Xu 19851

Value-Rich, Differently Organized: Measuring Value Expression and Organization in AI-Generated Debates
Maoyuan Li, Haoyuan Li, Zhongsheng Wang and Jiamou Liu 19871

LLM4LLM: Bridging Kernel Benchmarks and Real Deployment via Closed-Loop Agentic Optimization
Hui Zeng, Pengfei Yang, Yanxin Chen, Fusong Ju and Xinran Wei 19891

GIFT: Goal-Injected Fine-Tuning for Efficient Manipulation Policy Adaptation
Xiaoyuan Fang, Shuo Feng, Yuxuan Wang, Enhua Cheng, Peng Zhou and Piji Li 19903

Transporting Concepts Across Languages: Optimal Transport Toward English-Level Performance in Multilingual LLMs
Anh Trac Duc Dinh, Anh Duc Le and Tho Quan 19923

Two to Tango: Coupled Task-Reference Selection for Safe LLM Fine-tuning
Xinrui Chen, Jianhao Zhang, Ou Wu and Di Gao 19944

Culturally-Aware Alignment for Multimodal Models via Group Reward Optimization for Southeast Asia
Longyin Zhang, Shuo Sun, Yingxu He, Won Cheng Yi Lewis, Bowei Zou and AiTi Aw . . . 19969

When Response Variation Does Not Indicate Uncertainty: Hallucination Detection for Multi-Answer Queries
Hongliang Dai, Qiuyuan Li and Piji Li 19989

PRISM: Predictive Recomposition via Semantic Latent Decomposition for View-invariant Video Representation Learning
Youngchae Chee, Hosu Lee, Sungjune Park, Junho Kim and Yong Man Ro 20014

AlphaToken: Decoupling Adaptation and Stability for Path-Aware Response Token Valuation in LLM Post-Training
Qing Liu, Ou Wu and Yi Du 20028

Long Live the Librarian! A Persistent Search Sub-Agent for Energy-Efficient Multi-Agent Software Engineering Systems
Seunghyuk Cho, Sunghyun Choi, Jaeseung Heo, Youngbin Choi, Saemi Moon, MoonJeong Park and Dongwoo Kim 20055

State-Conditioned Visual Evidence Retrieval for Fine-Grained Perception in Document Vision-Language Models
Mingxu Chai, Chenyu Liu, Ziyu Shen, Jiazheng Zhang, Kaidi Zhang, Jihua Kang, Tao Gui and Qi Zhang 20078

R²Edit: Mastering Relation Replacement Editing via Target-Smoothed Forgetting and Self-Paced Learning
Xu Cao, Xinrui Chen, Qing Liu, Mengyang Li, Ou Wu and Yi Du 20096

Prefix-Adaptive Block Diffusion for Efficient Document Recognition
Mingxu Chai, Ziyu Shen, Chenyu Liu, Jihua Kang, Tao Gui and Qi Zhang 20124

Using LMs to Model the Effects of Context and Coreference during Sentence Comprehension
Kohei Kajikawa, Lin Ai, Tatsuki Kuribayashi and Ethan Gotlieb Wilcox 20139

Managing Evidence at Document Scale: A Survey of Multi-Page Visually Rich Document Understanding
Lewei Xu, Yihao Ding, Zihan Xu, Siwen Luo, Daochang Liu and Wei Liu 20159

EEMA: Experience-Evolving Meta Agents for AutoML
Kai Tian, Weizhi Wang, Kaikai Zhao, Che Jiang, Junlin Yang, Zhenzhao Yuan, Jincheng Zhong, Guoli Jia, Yuru Wang, Xinwei Long, Ning Ding, Bowen Zhou and Kaiyan Zhang 20188

Nullify: Null-Space Activation Steering for Training-Free LLM Unlearning
Wei Zhai, Xiang Liu, Qiang Huang, Rui Qian, Lemao Liu, Ziwei Li, Ziqi Wang, Zhitao Huang and Dejing Dou 20213

Support Vector Rubrics: Closing the Gap Between Self-Generated and Human Rubrics
Mengyuan Sun, Yu Li, Zhuohao Yu, Shikun Zhang and Wei Ye 20226

How Do Self-Referential LLM Loops Contract Political Diversity?
Wooyoung Kim and Wooju Kim 20246

TrustDLM: Benchmarking Trustworthiness in Diffusion Language Models
Yichuan Mo, Yukun Jiang, Yanbo Shi, Mingjie Li, Michael Backes, Yang Zhang and Yisen Wang 20264

MASCOT: Multi-Agent Socio-Collaborative Companion Systems
Yiyang Wang, Yiqiao Jin, Alex Cabral and Josiah Hester 20293

Who Counts as the People"? Measuring Contested Political Concepts from Political Speech using Masked Language Models
Hyeongjae Lee, Seonwoo Ahn, Sunwoo Kwak, June Jeon, Jisung Yoon and Taegyoon Kim 20314

Empathy Is Steerable but Multi-Axial: Mechanism Geometry and Persona Effects in LLMs
JuHeon Ha, Byounghan Lee, Yunseo Choi and Kyung-Ah Sohn 20326

INSIGHT: Structured Multimodal Knowledge Modeling for Vision-Guided Task-Oriented Dialogue
Xuelian Dong, Jiale Chen and Tianyong Hao 20344

OmniHalluc-L: Counterfactual Benchmarking and Modality-Perturbation Calibration for Long-Form Omni Hallucination
Zixuan Dong, Jiafu Tang, Zhide Lei, Zhe Cao, Zijie Zhang, Yanghai Wang, Shihao Li, Xiaodong Wang, Baoyun Peng and Jiaheng Liu 20361

Pak3H: Evaluating the Cost of Cultural Mismatch in LLM Alignment with a Human-Contextualized Urdu Benchmark
Abdullah Hashmat, Usman Naseem and Agha Ali Raza 20378

When Irrelevant Text Matters: Affine Margin Shifts in Multimodal Large Language Models
Yinfeng Wang, Zhiyuan Yao, Zheren Fu, Lei Zhang and Zhendong Mao 20395

HMPO: Hybrid Median-length Policy Optimization for Chain-of-Thought Compression
Minghui Zheng, Hongxu Chen, Huimin Ren, Hongsheng Xin, Xiaoyang Qu, Ze Wang, Shuling Yang, Ziyu Peng, Kaike Zhang, Zhou Pan and Kun Zhan 20418

NEAR: Neighborhood-Aware Context Reconstruction for Ultra-Long Document Reading Comprehension of LLMs
Han Li, Xuanang Chen, Ben He and Le Sun 20434

ScaffoldAgent: Utility-Guided Dynamic Outline Optimization for Open-Ended Deep Research
Zhibang Yang, Xinke Jiang, Yuzhen Xiao, Ruizhe Zhang, Yue Fang, Xinfei Wan, Zhengxing Song, Yuxuan Liu, Yuheng Huang, Junfeng Zhao, Yasha Wang and Xu Chu 20454

PALMs: Using Multi-Construct Grounded Rationales for Modeling Population Preferences in LLMs
Priyanka Dey, Brihi Joshi, Preyashi Poddar, Jieyu Zhao and Emilio Ferrara 20475

Decoding One Safety Trigger Token for Balancing Safety and Usability in Large Language Models
Haoran Gu, Handing Wang, Yi Mei, Mengjie Zhang and Yaochu Jin 20518

KG-RLVR: Reinforcing LLM Reasoning with Knowledge Graph-Derived Process Rewards
Aoran Gan, Guoxuan Zhang, Ya Li, Qiangbing Wang, Quan Liu, Qi Liu and Guoping Hu . 20538

PropUQ-MAS: Propagation-Aware Uncertainty Quantification for LLM Multi-Agent Systems
Yaokun Liu, Yifan Liu, Daniel Yue Zhang, Ruichen Yao, Zelin Li and Dong Wang 20555

ECHO: Replacing the Compliance Proxy with Response Harmfulness in Optimization-Based Jailbreak Attack
Yuanwei Wu, Lingdong Huang, Yunlong Mao, Kanye Ye Wang, Jingyu Hua and Sheng Zhong 20576

SEAR: Self-Judging Agentic Reward Models
Bingqian Li, Chuan Hao, Junyi Li, Ran Tao, Jian Yang and Xin Zhao 20591

SPADER: Step-wise Peer Advantage with Diversity-Aware Exploration Rewards for Multi-Answer Question Answering
Qiming Shi, Zhaolu Kang, Yunfan Zhou, Di Weng and Yingcai Wu 20604

Can LLMs Be Constrained to the Past? Improving Knowledge Cutoff through Recall-Based Prompting
Michiro Asai, Ailiang Lin, Yu Kishimoto, Takao Obi, Satoshi Kosugi, Kotaro Funakoshi and Manabu Okumura 20622

LongSocialBench: Do Long-Context LLMs Understand Online Discussion Threads?
Xinyi Liu, Rinat Khaziev, Dilek Hakkani-Tür and Tarek F. Abdelzaher 20638

Towards Holistic Knowledge Retrieval for Augmented Generation: Integrating Graph-Grounded and Query-Aligned Knowledge
Weidong Zhang, Zhe Wang, Kewen Wang, Haifeng Guo, Shuxuan Lian and Weizhi Du . . . 20661

HyperRAG-IHM: Incremental Hyperedge Maintenance for Hypergraph RAG under Corpus Expansion
Haoyu Chen, Jihua Zhu, Tingkai Jia, Shengwei Zhao, Lin Chen and Xiaoxuan Mu 20682

Cognitively Decoupled Meta-Induction for Out-of-Distribution Generalization
Yuqian Wang, Yuanzheng Wang, Chaojun Xiao, Ke Xu, Bingxiang He, Jinghan Tan, Zijun Chen, Lu Chen, Yi-Xin Huo and Maosong Sun 20694

ProbeTQA: Verifiable Table Question Answering via Multi-Round Executable Probing
Yan Liang, Huimin Ren, Xindi Zhang, Hongsheng Xin, Wenbin Huo, Yanan Guo, Jinpeng Zheng, Kaike Zhang, Kun Zhan and Jiani Hu 20715

Whose Norms? Disentangling Cultural and Personal Alignment in Large Language Models
Angana Borah, Isabelle Augenstein and Rada Mihalcea 20742

Entropy-Guided Token Dropout: Training Autoregressive Language Models with Limited Domain Data
Jiapeng Wang, Yiwen Hu, Yanzipeng Gao, Haoyu Wang, Shuo Wang, Hongyu Lu, Jiaxin Mao, Xin Zhao, Junyi Li and Xiao Zhang 20783

WebTraveler: A Multi-Agent Framework for Continuous Exploration in Website Traversal
Rongchang Li, Yang Shen, Gan Tang, Zizhao Zhang, Shuai Zhang, Haipang WU and Xi Li 20800

ScreenHaystack: Finding Blind Zones in GUI Grounding
Chenyue Li, Xiaoxiao Sun, Yubo Deng, Qinlin Zhao, Serena Yeung-Levy and Yuhui Zhang 20821

Can MLLMs Read the Room? A Multimodal Benchmark for Assessing Deception in Multi-Party Social Interactions
Caixin Kang, Yifei Huang, Liangyang Ouyang, Mingfang Zhang, Ruicong Liu and Yoichi Sato 20839

You Shouldn't Have Asked: A Pragmatics-Inspired Taxonomy for Evaluating LLM Refusals
Ruoxuan Li, Pinqiao Wang, Sheng Li and Cameron Robert Jones 20869

Guardrail Models Collapse under Adversarial Paraphrases
Yusuke Hirota, Shaona Ghosh, Yu-Chiang Frank Wang and Ryo Hachiuma 20895

From Triage to Discharge: A Survey of NLP Tasks, Methods, and Open Challenges in the Emergency Department
Dipankar Srirag, Aditya Joshi, Salil S. Kanhere and Padmanesan Narasimhan 20918

Decoding Diffusion Language Models with Foreseeing Movement
Yichuan Mo, Quan Chen, Mingjie Li, Zeming Wei and Yisen Wang 20943

MODE: Modality-Decomposed Expert-Level Mixed-Precision Quantization for MoE Multimodal LLMs
Yuanteng Chen, Nanxin Zeng, Peisong Wang, Zhilei Liu, Yuantian Shao, Shiqiang Lang, Tao Liu, Chuangyi Li, Qinghao Hu, Gang Li, Jing Liu and Jian Cheng 20959

Self-Evolving Optimization of Agentic Systems for Machine Translation
Zhan Shen, Xiaotian Wang, Jason Naradowsky and Yusuke Miyao 20978

Less Is Moral: A CHARMing Framework for Moral Foundations Detection in Endorsement Behaviour
Huixiang Fu and Marian-Andrei Rizoiu 21003

Aggregation-Aware Synthetic Text Generation Against Authorship Re-Identification
Qian Ma, Anna Cinzia Squicciarini and Sarah Rajtmajer 21026

Tracing Stereotypes from Representation to Output in Multilingual LLMs
Ariun-Erdene Tumurchuluun, Yusser Al Ghussin, Pinzhen Chen, Josef Van Genabith and Koel Dutta Chowdhury 21057

PAGE: Progress-Aware Reinforcement Learning with Hierarchical Advantage for Long-Horizon GUI Agents
Qingliang Zhang, Yongqian Sun, Yuanhong Cai, Xiaohui Nie, JiaJu Wang, Yu Luo, Jiamin Jiang, Zhekang Chen, Yaozheng Fang, Meng Wang, Chaobo Xian, Peng Li and Dan Pei 21075

Not All Thoughts are Generated Equal: Efficient LLM Reasoning via Synergizing-Oriented Multi-Turn Reinforcement Learning
Yansong Ning, Wei Li, Jun Fang, Naiqiang Tan and Hao Liu 21102

CODE: A Contradiction-Based Overthinking Attack Framework for Prolonging Retrieval-Augmented Generation
Xiaolei Zhang, Xiaojun Jia, Liquan Chen and Songze Li 21127

Large Language Models in Resolving Contextual Knowledge Conflicts
Xinye Yang, Zhenyang Liu, Ruisi Li and Yuanyuan Lei 21142

OmniACBench: A Benchmark for Evaluating Context-Grounded Acoustic Control in Omni-Modal Models
Seunghee Kim, Bumkyu Park, Kyudan Jung, Joosung Lee, Soyoon Kim, Jeonghoon Kim, Taeuk Kim and Hwiyeol Jo 21170

Analyzing and Mitigating Cross-Lingual Degradation in Multilingual Medical VQA
Jingbo Wang, Sendong Zhao, Haochun Wang, Bing Qin and Ting Liu 21191

Is Next-Chunk Reasoning RL Really Better than SFT? Revisiting Training Strategies under no-CoT Data
Yinhao Tang, Youqing Fang, Yanan Sun, Jiangning Liu, Ziyi Wang, Xun Zhao, Weiming Zhang, Bin Liu, Kuikun Liu, Wenwei Zhang and Kai Chen 21209

DataSight: Learning to Predict Before Executing for Data Analysis Agents
Shangke Lyu, Zhe Zheng, Linjuan Wu, Yan Yu, Xuan Zhang, Jin Ma, Yongliang Shen and Weiming Lu 21229

Interpreting and Steering for Safe and Correct Code Generation
Hao Yan and Ziyu Yao 21252

OmniDenoise: Diffusion MultiModal LLMs for Omni-Modal Understanding and Prediction
Xianfeng Wu, Yuhang Chen, Jie Peng, Jinhao Duan and Tianlong Chen 21273

AgentHallu: Benchmarking Automated Hallucination Attribution of LLM-based Agents
Xuannan Liu, Xiao Yang, Zekun Li, Pei Pei Li and Ran He 21290

Subjective and Hierarchical Annotation of Text Difficulty in Japanese
Daisuke Maekawa, Kazumasa Omura, Sora Tarumoto, Shotaro Ishihara and Tomoyuki Kajiwara 21316

Latent Ordinal Evidence, Misaligned Outputs: Inference-Time Ordinal Lens Alignment for Multimodal LLMs
Haiming Li, Yingsheng Liu, Jingmin Zhu, Siyuan Yan, Xieji Li, Jiajun Sun, Zongyuan Ge and Zhen yu 21328

Fake It Till You Make It: Training Deep-Search Agents in Synthetic Search Environments
Yibo Xue, Pin Zheng, Chengyu Wang, Yukun Kang, Yuanjie Lyu, Ming Wang, Jian Peng, Jun Huang and Changliang Xu 21347

LLM as GNN: Graph Vocabulary Learning for Text-Attributed Graph Foundation Models
Xi Zhu, Haochen Xue, Ziwei Zhao, Wujiang Xu, Ching-Yuen Huang, Minghao Guo, Qifan Wang, Kaixiong Zhou, Imran Razzak and Yongfeng Zhang 21373

Illusion of Alignment: Detecting Hidden Disagreement in Collaborative Dialogue
Kaiming Liu, Fuwen Luo, Ziyue Wang, Jinrui Ju, Yuxuan Liu, Xuanyu Lei, Yunghwei Lai, Peng Li and Yang Liu 21395

From 'May' to 'Is': Certainty Distortion in Language Model Rewriting
Catarina G Belém, Shang Wu, Randol H. Yao, Mark Steyvers, Sameer Singh and Padhraic Smyth 21426

Beyond the Payload: How User Invocation Shapes Coding Agent Vulnerability to Repository Poisoning
Fukang Zhu, Binbin Zhao, Ruixiao Lin, Ping He, Tianyu Du and Shouling Ji 21465

Cross-Lingual Representation Alignment by Token-Level Optimal Transport in a Language-Agnostic Space
Taisei Yamamoto, Ryoma Kumon, Danushka Bollegala and Hitomi Yanaka 21495

SCORE: Specificity, Context Utilization, Robustness, and Relevance for Reference-Free LLM Evaluation
Homaira Huda Shomee, Rochana Chaturvedi, Yangxinyu Xie and Tanwi Mallick 21515

See Beyond Words: Joint Visual–Language Representations Improve Brain Encoding of Language-Only Stimuli
Yongpeng Zhang, Congchi Yin, Runze Xia, Xuyun Wen and Piji Li 21543

SOPBench: Evaluating Language Agents at Following Standard Operating Procedures and Constraints
Zekun Li, Shinda Huang, Jiangtian Wang, Nathan Zhang, Antonis Antoniades, Wenyue Hua, Kaijie Zhu, Sirui Zeng, Chi Wang, William Yang Wang and Xifeng Yan 21561

ConDistill: Asymmetric Context Self-Distillation for Radiology Report Generation
Ziming Song, Mingyang Shen and Chi Man Vong 21617

Your Agent, Their Asset: A Real-World Safety Analysis of OpenClaw
Zijun Wang, Haoqin Tu, Letian Zhang, Hardy Chen, Juncheng Wu, Xiangyan Liu, Zhenlong Yuan, Tianyu Pang, Qizhe Xie, Fengze Liu, Zeyu Zheng, Huaxiu Yao, Yuyin Zhou and Cihang Xie . . . 21631

Can Heterogeneous Language Models Be Fused?
Shilian Chen, Jie Zhou, Qin Chen, Wen Wu, Xin Li, Feng Qi and Liang He 21650

MER-Bench: Benchmarking Multimodal Emotional Reappraisal for Image-Text Memes
Yiqi Nie, Fei Wang, Junjie Chen, Kun Li, Yudi Cai, Dan Guo, Chenglong Li and Meng Wang 21670

Beyond Over-Simplification and Over-Personalization: Lexical Complexity Prediction for Leveled Reading
Huidan Xu and Ying Liu 21689

PISanitizer: Preventing Prompt Injection to Long-Context LLMs via Prompt Sanitization
Runpeng Geng, Yanting Wang, Chenlong Yin, Minhao Cheng, Ying Chen and Jinyuan Jia 21706

SyRHM: Symbolic-Language-Enhanced Reasoning with Associative Retrieval for Zero-shot Harmful Meme Detection
Hanling Wang, Chenlong Wei, Yingjuan Li, Di Wu, Yuchao Zhang, Xiaohui Zhu and Yao Zhu 21730

A Benchmark Construction and Evaluation Framework for Specialist Domains: Case Study on Defense-Related Documents
Bao Gia Doan, Aditya Joshi, Pantelis Elinas, Aarya Bodhankar, Oscar Leslie, Tom Marchant and Flora D. Salim 21747

SLIT: Execution-Grounded Small Language Models for Issue-Reproducing Test Generation
Runlin Liu, Zhe Zhang, Zeren Chen, Xiang Gao, Lu Sheng, Ping Yang, Mu Li and Hailong Sun 21776

Comparing Linear Probes with Mahalanobis Cosine Similarity
Zhuofan Ying, Peter Hase and Nikolaus Kriegeskorte 21793

Cross-Lingual Alignment Without Joint Training: Do Monolingual Language Models Converge on Universal Representations?
Ej Zhou, Suchir Salhan, Catherine Arnett and Anna Korhonen 21811

SkillHarm: Lifecycle-Aware Skill-Based Attacks via Automated Construction
Yuting Ning, Zhehao Zhang, Yash Kumar Lal, Boyu Gou, Junyi Li, Weitong Ruan, Chentao Ye, Rahul Gupta, Diyi Yang, Yu Su and Huan Sun 21838

VideoAgent: All-in-One Framework for Video Understanding and Editing
Hengji Zhou, Lingxuan Huang, Jian Wang, Bing Zhou, Si Wu, Lianghao Xia and Chao Huang 21860

Heard or Read? Separating Text Shortcuts from Audio Perception in Music Multiple Choice
Xinyang Wu and Andrew Horner 21907

MemeScope: Probing and Causally Localizing Pragmatic Inference in LVLM Meme Understanding
Tongguan Wang, You Wu, Yuling Zhou, Siyuan Chen, Junkai Li, Feiyue Xue, Wei Chen, Hui Liu, Xiqiao Ba and Ying Sha 21923

Do LLMs Understand Personality? Rethinking Persona Fidelity Evaluation through Structured Behavioral Inference
Mengfan Li, Zesheng Wei, Xuanhua Shi and Yang Deng 21946

BLADE: Bilevel Low-rank Augmented-Lagrangian Erasure for LLM Unlearning
Md Toufikuzzaman, Ahmad Mousavi and Dongwon Lee 21966

Towards Agentic Document-Grounded Information Seeking
Qintong Zhang, Xinjie Lv, Jialong Wu, Baixuan Li, Zhengwei Tao, Guochen Yan, Huanyao Zhang, Bin Wang, Jiahao Xu, Haitao Mi and Wentao Zhang 21988

Pattern Over-Generalization of Knowledge Graph Embedding
Junsik Kim and Kangil Kim 22014

Risk-conditioned Fine-tuning of Large Language Models
Zixuan Liu, Fangzheng WU, Brian Summa and Zizhan Zheng 22036

ReverseEOL: Improving Training-free Text Embeddings via Text Reversal in Decoder-only LLMs
Ailiang Lin, Zhuoyun Li, Yusong Wang, Keyu Mao, Kotaro Funakoshi and Manabu Okumura 22077

Beyond the Final Layer: Intermediate Representations for Better Multilingual Calibration in Large Language Models
Ej Zhou, Caiqi Zhang, Tiancheng Hu, Chengzu Li, Nigel Collier, Ivan Vulić and Anna Korhonen 22097

Beyond Blind Compliance: Benchmarking Task Verification in OCR Reasoning
Yue Zhou, Yuan Wu and Yi Chang 22122

GeoSID: Quantizer-Aware Geometric Shaping for Semantic IDs in Generative Recommendation
Xinrong He, Chenghao Wu, Lianrong Chen, Xuming Hu and Menglin Yang 22156

EgoTools: Towards Tool-Centric Reasoning in Real-World Egocentric Videos
Shulin Tian, Junsu Kim, Shuai Liu, Hao Li, Yujiao Shen, Sihan Li, Zhe Yang, Yeongon Kim, Runmao Yao, Yuhao Dong, Fangzhou Hong, Antonino Furnari, Jingkang Yang, Hongyuan Zhu and Ziwei Liu 22169

MAIGO: Mitigating Lost-in-Conversation with History-Cleaned On-Policy Self-Distillation
Haoyu Zheng, Yun Zhu, Shu Yuan, Shangming Chen, Qing Wang, Wenqiao Zhang, Jun Xiao and Yueting Zhuang 22191

Legal Rule Induction: Towards Generalizable Principle Discovery from Analogous Judicial Precedents
Wei Fan, Tianshi Zheng, Yiran HU, Zheye Deng, Weiqi Wang, Baixuan Xu, Chunyang Li, Haoran Li, Weixing Shen and Yangqiu Song 22212

RedEdit: Agentic Red-Teaming of Image Safety Classifiers via MCTS-Guided Photo-Editing
Weilin Lin, Ziqi Lin, Zhenxing Zhou, Jianze Li, Tong Zhang, Hui Xiong and Li Liu 22241

RegionDoc-R1: Reinforcing Semantic Layout-Aware Reasoning for Document Understanding
Chao Wang, Hehe Fan, Huichen Yang, Zhengdong Hu, Sarvnaz Karimi, Lina Yao and Yi Yang 22260

Does Understanding Inform Generation in Unified Multimodal Models? From Analysis to Path Forward
Yuwei Niu, Weiyang Jin, Jiaqi Liao, Chaoran Feng, Peng Jin, Bin Lin, Zongjian Li, Bin Zhu, Weihao Yu and Li Yuan 22283

AIVE: Active Imagined View Exploration for Visual Spatial Reasoning
Zhenming Wu, Li Li, Song Yu, Wenwen Zhao, Zhisheng Yang and Shiyu Zhu 22304

TRACE: Tree-Structured Reasoning for Multi-Table Multi-Hop Fact-Checking
Senyi Zhang, Dongwon Lee and Delvin Ce Zhang 22324

Evo-Harness: Context-to-Harness Skill Compilation for Self-Evolving Agents
Tianxin Wei, Zhan Shi, Minhua Lin, Bing He, Zewen Liu, Yisi Sang, Yuanchen Bei, Xuying Ning, Jiaru Zou, Ting-Wei Li, Xiao Lin, Yanjun Zhao, Chi Wang, Benoit Dumoulin, Dakuo Wang, Jingrui He and Hanqing Lu 22343

DSAEval: Evaluating Data Science Agents on a Wide Range of Real-World Data Science Problems
Maojun Sun, Yifei Xie, Yue Wu, Ruijian Han, Binyan Jiang, Defeng Sun, Yancheng Yuan and Jian Huang 22359

Sufficiency-Aware Adversarial Evidence Masking for Textual Emotion-Cause Pair Extraction in Conversations
Mengqiang Hu, Qian Chen, Xin Guo and Suge Wang 22394

Tokenization as Structural Inductive Bias: Grapheme-Constrained Subwords for Multilingual Modeling and Vocabulary Surgery
Arush Singhal, Jatin Bedi and Jasmeet Singh 22414

Generation-Score Consistency: Pushing LLMs to Generate Semantically Valid Reasoning Scores for Misinformation Detection
Yidu Chen, Bo Ma, Yating Yang, Xi Zhou, Zhen Wang, Ahtamjan Ahmat, Dilxat Abdureyim and Azmat Anwar 22435

Sell More, Play Less: Benchmarking LLM Realistic Selling Skill
Xuanbo Su, Wenhao Hu, Le Zhan, Yuting Xie, Kailin Lyu, Kaijie Chen, Ziwei Li, Yeqiang Wang, Haibo Su, Yunzhang Chen and Ling Huang 22456

ParaRecover: A Process-Level Benchmark for Error Localization and Recovery in Parallel Tool-Use Agents
Bowen Guan, Zhentao Yin and Yanming Shen 22479

CultureConverse: A Multilingual Multi-turn Simulation Harness for Culturally Grounded Assistance in East and Southeast Asia
Bryan Chen Zhengyu Tan, Weihua Zheng, Thong T. Doan, Bich Ngoc Doan, Jia Wang Peh, Xiaoyuan Yi, Jing Yao, Xing Xie, Nancy F. Chen, Zhengyuan Liu, JinYeong Bak, Wafi Shamdi, Soo Kai Chie, Liew Yu Siong, Aina Azyyati Binti Mohamad Rezal, Lew Yan Yan Vanessa, Huadan Wu, Dylan

Raharja, Nadya Yuki Wangsajaya, Akane Fukushige, Kazushi Kato, Koji Inoue, Tatsuya Kawahara, Jaehyung Seo, Dongjun Kim, Seungyoon Lee, Zi Haur Pang, Rui Yang Tan, Charibeth Ko Cheng, Maria Regina Justina Estuar, Jann Railey Montalan, Pham Minh Duc and Roy Ka-Wei Lee 22497

VTOS: Learning to Orchestrate Vision Tools by Co-Searching Solutions and Observers
Jinchao Ge, Lingqiao Liu, Shuwen Zhao and Lei Wang 22551

IndoDialect: A Multi-Dialect Conversational Dataset for Indonesian Local Languages
Salsabila Zahirah Pranida, Thamar Solorio and Fajri Koto 22570

MTIMEFACT: Improving Large Language Models for Multi-Hop Temporal Fact-Checking
Xinwei Yang, Ruiling Guo, Chen Huang, Wenqiang Lei and See-Kiong Ng 22592

When No Answer Is Correct: Diagnosing Absent Answer Detection for MLLMs in Video Understanding
Yiheng Wang, Yueqian Lin, Lichen Zhu, Yudong Liu, Hai Helen Li and Yiran Chen 22615

Semantic Head Specialization Guides Hybrid ViT Attention for Multimodal LLMs
Chenhong He, Lei Li, Shicheng Li, Hanglong Lv, Lingpeng Kong, Qi Liu, Tong Yang and Shuhuai Ren 22626

M^2G-Eval: Enhancing and Evaluating Multi-granularity Multilingual Code Generation
Fanglin Xu, Wei Zhang, Jian Yang, Guo Chen, Aishan Liu, Zhoujun Li, Xianglong Liu and Bryan Dai 22648

SlimKV: Joint Token-Feature KV Cache Compression with Reconstruction-Free Beacon Attention
Zihan Teng, Jiayu Zhao, Wentao Ren, Minhao Fan, Tianrui Ma, Song Chen and Weichen Liu 22662

OceanGym: Evaluating Language-Grounded Embodied Agents in Underwater Environments
Yida Xue, Mingjun Mao, Xiangyuan Ru, Yuqi Zhu, Baochang Ren, Shuofei Qiao, Mengru Wang, Shumin Deng, Xinyu An, Ningyu Zhang, Guozhou Zheng, Ying Chen and Huajun Chen 22678

CORTEX: High-Quality Cross-Domain Organization of Web-Scale Corpora through Ontological Corpus Graph
Chengtao Gan, Xiaoke Guo, Yushan Zhu, Zhaoyan Gong, Zhiqiang Liu, Songze Li, Huajun Chen and Wen Zhang 22700

TR-ICRL: Test-Time Rethinking for In-Context Reinforcement Learning
Wenxuan Jiang, Yuxin Zuo, Zijian Zhang, Xuecheng Wu, Zining Fan, Wenxuan Liu, Li Chen, Xiaoyu Li, Xuezhi Cao, Xiaolong Jin and Ninghao Liu 22725

ResMerge: Residual-based Spectral Merging of Large Language Models
Yandu Sun, ZhiYan Hou, Hongyan An, Weizhen Wang, Haokai Ma, Yuheng Jia, Junfeng Fang, Haiyun Guo and Jinqiao Wang 22741

Qwen-Image-Bench: From Generation to Creation in Text-to-Image Evaluation
Niantong Li, Guangzheng Hu, Weixu Qiao, Ying Ba, Qichen Hong, Shijun Shen, Jinlin Wang, Fan Zhou, Jianye Kang, Xin Shang, Ziyi HE, Wei Wang, Dalin Li, Jiahao Li, Jie Zhang, Kaiyuan Gao, Kun Yan, Lihan Jiang, Ningyuan Tang, Shengming Yin, Tianhe Wu, Xiao Xu, Xiaoyue Chen, Yuxiang Chen, Yan Shu, Yanran Zhang, Yilei Chen, Yixian Xu, Zekai Zhang, Zhendong Wang, Zihao Liu, Zikai Zhou, Hongzhu Shi, Yi Wang, Bing Zhao, hu Wei, Lin Qu and Chenfei Wu 22760

Causal Interventions on Continuous Variables: A Case Study on Verb Bias in Steering Vectors for In-Context Learning
Zhenghao Herbert Zhou, R. Thomas McCoy and Robert Frank 22788

Robust Failure, Conservative Repair: Textual Knowledge Distillation from Cross-Model Failures
Andrew Ren, Haokun Liu and Chenhao Tan 22806

ECT-CLIP: Resource-Efficient CT Vision-Language Pretraining via Negative Scaling
Hanwen Zhang, Yao Liu, Jiaye Yang, Die Dai, Qiao Liu, Yutong Xie and Peng Wang 22822

S3Mem: Structured Spatiotemporal Scene-Event Memory for Long-Horizon Interactive Question Answering
Encheng Su, Jianyu Wu, Jinouwen Zhang, Qiucheng Yu, Chen Tang, Pengze Li, Lintao Wang, Aoran Wang, Xinzhu Ma, Shixiang Tang, Yizhou Wang and Houqiang Li 22835

When New Generators Arrive: Lifelong Machine-Generated Text Attribution via Ridge Feature Transfer
Zhen Sun, Yifan Liao, Zhicong Huang, Jiaheng Wei, Cheng Hong, Yutao Yue and Xinlei He 22853

AnalysisBank: An Expert Analysis Pattern Library for Financial Report Generation
Yajing Yang, Yunshan Ma, Kelvin J.L. Koa and Min-Yen Kan 22870

Decoupling Turn-Taking from Semantics: A Decoupled Data Approach for Finite-State-Machine-Based Full-Duplex Dialogue
Yihang Li and Chenhui Chu 22893

Controllable Text-Attributed Graph Generation via LLM-Based Substructure Modeling
Ruijun Mao, Jiaxuan Chen, Boyan Li, Zixing Zhang, Yuhang Zhang and Fangxin Wang . . . 22911

PRISM: Refracting Large-Batch Model Editing through Jacobian-Inferred FFN Subspaces
Ziyue Zhai, Qiyao Liu, Hao Song, Heng Yu, Guozhi Lin, Yulai Yang and Xuan Zhang 22931

Small Changes, Big Impact: Demographic Bias in LLM-Based Hiring Through Subtle Sociocultural Markers in Anonymised Resumes
Bryan Chen Zhengyu Tan, Shaun Khoo, Bich Ngoc Doan, Zhengyuan Liu, Nancy F. Chen and Roy Ka-Wei Lee 22950

How Well Does Mathematical Reasoning Transfer Across Languages in English-Centric LLMs?
Wen Yang, Junhong Wu, Chong Li, Chengqing Zong and Jiajun Zhang 22983

How Well Can Modern LLMs Act as Agent Cores in Radiology Environments?
Qiaoyu Zheng, Chaoyi Wu, Weike Zhao, Pengcheng Qiu, Lisong Dai, Ya Zhang, Yanfeng Wang and Weidi Xie 23010

Solve the Missing First Step: Can VLMs Standardize Raw Heterogeneous Medical Data?
Xin Chen, Dongliang Xu, Cunhao Zhu, Xudong Luo, Haoyang Lyu, Xiaoxiao Sun, Serena Yeung-Levy and Yue Yao 23041

When Modality Gap Reduction Fails: Prediction-Level Hubness in CLIP
Shota Sato, Hajime Kiyama, Tosho Hirasawa and Mamoru Komachi 23060

SafeRx-Agent: A Knowledge-Grounded Multi-Agent Framework for Safe and Explainable Medication Recommendation
Xinyu Wang, Hanwei Wu, Zhenghan Tai, Sicheng Lyu, Qincheng Lu, Ziyu Zhao, Jijun Chi, Jingrui Tian, Xiao-Wen Chang and Ziyang Song 23084

InfoQuant: Shaping Activation Distributions for Low-Bit LLM Quantization
Ke Li, Dong AN, Xiaoling Zang, Can Ye, Liang Xie, Qibo Qiu, Chen Shen, Xiaofei He and Wenxiao Wang 23113

Trust Cascade through Provenance Erasure: How Multi-Agent Delegation Breaks Single-Agent Safety
Yuxin Zhou, Xiao Ding, Tianle Chang, Ye He, Wenbin Zhang, ShangZhan Li, Yijia Meng, Wenyu Zang, Bibo Cai and Qi Shi 23130

TaxoAgent: Utility-Aligned Policy Optimization for Hierarchical Semantic Classification
Zhiming Lin, Yihao Zhong, Liwei Hou, Saisai Hu, ShaoYan He, Chen Rongru, Xuan Yang and Siyue Liu 23149

CoDAR: Continuous Diffusion Language Models are More Powerful Than You Think
Junzhe Shen, Jieru Zhao, Ziwei He and Zhouhan Lin 23172

CASTLE: A Comprehensive Benchmark for Evaluating Student-Tailored Personalized Safety in Large Language Models
Rui Jia, Ruiyi Lan, Fengrui Liu, Zhongxiang Dai, Bo Jiang, Jing Shao, Jingyuan Chen, Guandong Xu, Fei Wu and Min Zhang 23188

MobileGym: A Verifiable and Highly Parallel Simulation Platform for Mobile GUI Agent Research
Dingbang Wu, Rui Hao, Haiyang Wang, Shuzhe Wu, Han Xiao, Zhenghong Li, Bojiang Zhou, Zheng Ju, Zichen Liu, Lue Fan and Zhaoxiang Zhang 23223

NEST: A Benchmark for Explicit Structural Parsing on Native Empty Templates
Chen Peng, Tianci He, Hongye Chen, Jiayi Huang, Kangdi Hao and Yiping Song 23248

How Coding Agents Fail Their Users: A Large-Scale Analysis of Developer-Agent Misalignment in 20,574 Real-World Sessions
Ningzhi Tang, Chaoran Chen, Gelei Xu, Yiyu Shi, Yu Huang, Collin McMillan, Tao Dong and Toby Jia-Jun Li 23264

HALT: Evaluator-Guided Clause-Level Pruning and Early Halting for Lower-Cost Multi-Agent Systems
Zongwei Tang, Yewen Li, Bo Liu, Zezong Zhang, Cai Yiqing, Yiwei Guo, Houjie Qiu and Xingkong Ma 23283

MosaicAudio: Multi-Teacher On-Policy Distillation for Controllable Video-to-Audio Generation
Huadai Liu, Peiwen Sun, Qian Chen, Wen Wang, Bin Ma, Xiangang Li and Wei Xue 23308

From Retrieval to Reconstruction: Constructing Evolvable Cognitive Memory for Long-term Dialogue
Zirui Liao, ZhengXian Wu, Zhuohong Chen, Yunyao Yu, Xiaoyu Liu, Yifan Xu and Haoqian Wang 23325

Domino: Decoupling Causal Modeling from Autoregressive Drafting in Speculative Decoding
Jianuo Huang, Yaojie Zhang, Qituan Zhang, Hao Lin, Hanlin xu and Linfeng Zhang 23346

EgoArgus: Benchmarking VLMs as Situational Assistants for Modality-Grounded User Supports
Yu-Chien Tang, Yu-Hsiang Liu and An-Zi Yen 23357

Dude: A Dual-Detection Multi-Agent System for Paper-Code Discrepancy Detection
Weijie Liu, Running Zhao, Wenhao Yuan, Jinfeng Xu, Zhanfeng Xu, Xiaoxi Zhang and Edith Cheuk-Han Ngai 23371

DeltaMem: Towards Agentic Memory Management via Reinforcement Learning
Qi Zhang, Shen Huang, Liuchu, Shouqing Yang, Xiaomeng Hu, Pengjun Xie, Haobo Wang and Junbo Zhao 23390

From Latent Empathy to Structured Training Signals for Emotional Support Dialogue
Linran Xu and Xiaomei Nie 23418

DiscoTrack: A Multilingual LLM Benchmark for Discourse Tracking
Lanni Bu, Lauren Levine and Amir Zeldes 23439

RoboSPA: Can VLA Models Go Beyond Simple Scenes and Short-Horizon Tasks?
Zhenxuan Fan, Bo Zhang, Yutong Lin, Yuqian Yuan, Juekai Lin, Liang Liang, Zhuoyi Huang, Wenqiao Zhang, Juncheng Li, Siliang Tang, Jun Xiao and Yueting Zhuang 23457

Import What You Need: Learning When and How to Augment EHR Graphs with External Knowledge
Chen Chen, Mohsen Nayebi Kerdabadi, Dongjie Wang, Mei Liu and Zijun Yao 23493

WorldLines: Benchmarking and Modeling Long-Horizon Stateful Embodied Agents
Yehang Zhang, Jianchong Su, Haojian Huang, Yifan Chang, Tianhao Zhou, Xinli Xu, Yingjie Xu, Yinchuan Li, Zexi Li and Ying-Cong Chen 23512

LifeMem: Enabling Lifelong Experience Reuse for LLM Agents
Yuli Qiu, Yutong Li, Wei Su, Zeming Liu, Wanxiang Che, Heyan Huang, Haifeng Wang and Yuhang Guo 23537

WRAP++: Web Discovery Amplified Pretraining
Jiang Zhou, Yunhao Wang, Xing W, TingHao YU and Feng Zhang 23570

Benchmarking Aggregation Queries over Unstructured Text
Haojia Zhu, Jiahui Jin, Haoyu Li, Qinyuan Xu, Yuxi Liu, Zihan Zhou, Hanchen Qiu, Jiaoyan Chen and Jinghui Zhang 23591

Beyond Dense States: Sparse Transcoders as Causally Testable Operators for LLM Latent Reasoning
Yadong Wang, Haodong Chen, Yu Tian, Chuanxing Geng, Dong Liang and Xiang Chen . . 23616

Do GUI Agents Know When Not to Act? Enabling Conflict-Aware Termination for Multimodal GUI Agents
Zhaoyuan Huang, Tianjie Ju, Pengzhou Cheng, Zheng Wu, Yansi Li, Chuanbiao Song, Jun Lan, Huijia Zhu, Weiqiang Wang and Zhuosheng Zhang 23630

Beneath the Diff: Diagnosing and Mitigating Algorithmic Mode Collapse in Code-Level Autonomous Research Loops
Bowei He, Weixu Zhang, Yili Jin and Xue Liu 23652

E2E-VisRAG: End-to-End Visual Retrieval-Augmented Generation with Differentiable Retrieval
Zehua Wang, Zhaojin Zhang, Xiaolong Weng, Yanzhi Chen, Ying Xiong and Buzhou Tang23670

Chiaroscuro for Emotions: A Contrastive Emotion Benchmark Grounded in Appraisal Theory
Divyesh Bommana, Mohammad Saim and Tianyu Jiang 23684

Bulbul: A Dataset for Dialectal Arabic Speech Recognition
Ahmed Ashraf, Aisha Alansari, Fadel Al Abbas, Nada Almarwani, Samah Aloufi, Saad Ezzini, Maged S. Al-shaibani, Doaa Dalaq, AbdelRahim A. Elmadany, Muhammad Abdul-Mageed, Mohamed Mehdi Trigui, Dania Refai, Layan Refai, Mohamed Akrout, Mustafa Jarrar, Wasfi G. Al-Khatib, Alaa Dalaq, Darin El-Nakla, Samir Abdaljalil, Abdulrahman Al-Fakih, Nour El Imane Zeghib, Moussa Redah, Salmane Chafik, Mohamed El-Attar, Rima Grati, Sarah Kohail, Malak Alkhorasani, Khadijah Al Safwan, Ismail Mohmmed Mudhaffar, Ali Altam, Ahmed Al-Shaikh, Adnan Mahyoub Ahmed Saeed and Hamzah Luqman 23702

TRIAD: A Three-Axis Framework for Knowledge Conflict Resolution in RAG
Yizhou Jin, Xudong Li, Dandan Song, Yue Yang, Jun Yang, Yuhang Tian, Huipeng Ma, Chenhao Li and Luan Zhang 23721

TSHMem: A Temporal-Semantic Hierarchical Memory Graph for Long-Term Conversational Agents
Feiyang Xu, Wenhao Yu, Li Xiang, Quanyu Dai and Irwin King 23747

Gated Contrastive Decoding via Local Candidate Geometry for Audio-Visual Speech Recognition
Da-Hee Yang and Joon-Hyuk Chang 23767

Reasoning and Tool-use Compete in Agentic RL:From Quantifying Interference to Disentangled Tuning
Yu Li, Mingyang Yi, Xiuyu Li, Ju Fan, Fuxin Jiang, Binbin Chen, Peng Li, Jie Song and Tieying Zhang 23779

Counterfactual Constraint-Conditioned On-Policy Distillation for Multi-Constraint Instruction Following
Yanzhao Zheng, Yuanqiang Yu, Tianze Xu, Chao Ma, Zhentao Zhang, Jihuai Zhu, Baohua Dong, Hangcheng Zhu and Ruohui Huang 23799

VisEditBench: Can Vision-Language Models Edit Visualization Code from Multimodal Feedback?
Mizanur Rahman, Arshia Azimlu, Shadikur Rahman, Md Tahmid Rahman Laskar, Amran Bhuiyan, Shafiq Joty and Enamul Hoque Prince 23828

Argument Collapse: LLMs Flatten Long-Form Public Debate
Yekyung Kim, Yapei Chang, Chau Minh Pham and Mohit Iyyer 23847

BenchPreS: A Benchmark for Context-Aware Personalized Preference Selectivity of Persistent-Memory LLMs
Sangyeon Yoon, Hyesoo Hong, Wonje Jeung, Yongil Kim, Wooseok Seo, Heuiyeen Yeen, Albert No and Sunkyoung Kim 23894

Same Chart, Different Story: Bias in Vision-Language Chart Interpretation
Mizanur Rahman, Huan Wu, Arash Asgari, Enamul Hoque Prince and Laleh Seyyed-Kalantari 23909

When Do Memories Break? Chain-of-Stage Diagnosis for LLM Memory Systems
Chenchen Wang, Yujie Fu, Zhiyu Zheng, Shiyu Qin, Zhao Cao and Zheng Liu 23931

MemDefrag: Latent Memory Defragmentation for Large Language Models
Ruiyi Yan, Zhuoyuan Mao and Steven Y. Guo 23952

The Harder Text Embedding Benchmark (HTEB): Beyond One-dimensional Static Robustness
Manuel Frank and Haithem Afli 23973

HyperStyler: Low-resource Authorship Style Transfer via Context-aware Style Navigation and Hypernetworks
Jongkyung Shin, Minguk Jeon, Chan Woo Park and Chiehyeon Lim 24006

DiffuTester: Accelerating Unit Test Generation for Diffusion LLMs via Mining Structural Pattern
Lekang Yang, Yuetong Liu, Yitong Zhang and Jia Li 24029

Routing-Aware Safety Alignment for Mixture-of-Experts Models
Jiacheng Liang, Yuhui Wang, Tanqiu Jiang and Ting Wang 24049

SoftGene: Protein Language Model-Enhanced Soft Prompting for Interpretable Gene Set Annotation
Drew Ross, Arya Hadizadeh Moghaddam, Dongjie Wang, Xiaoyu Zhang and Zijun Yao . . 24071

The Selection Effect: Why Compact LLM Verifiers for Emotion Classification Fail Behind Encoder Confidence Gates
Dmitry Tikhomirov and Andrey Savchenko 24095

Dense Process Supervision for Search Agents via Fact Utility Estimation
Rongzhi Zhu, Xiangyu Liu, Yi Liu, Shuo Zhang, Ruirui Zhang, Rui Wu, Tao Jiang, Zequn Sun, Wenhao Xu and Wei Hu 24115

Anchoring What Matters: A Dual-Level Learning Framework for Visually-Grounded Multimodal Reasoning
Xinxin Song, Siyuan Li, Tingxiong Xiao and Jinli Suo 24131

Privacy-Preserving Text Sanitization for Distributed Agents Collaboration via Disentangled Representations
Xuan Liu, Hefeng Zhou, Sicheng Chen, Chao Yang, Xingcheng Xu, Jingjing Qu, Jiong Lou, Jie LI and Xia Hu 24148

Predict, Don't Iterate: Efficient Adaptive-Length Infilling for Diffusion Language Models
Haobo Xu, Sirui Chen, Yuanchen Bei, Lingjie Chen, Yuchen Yan, Dongqi Fu, Jingrui He and Hanghang Tong 24171

CAS: Conformalized Agentic Search via Adaptive Retrieval and Policy Weighting
Zixi Zhu, Jiayuan Su, Jian Zhang, Yu Lin and Hongwei Wang 24191

GRAFE: Graph-Reasoned Automated Feedback and Evaluation for Structure-Aware Essay Assessment
Hanyao Wei, Jun Zhong, Guang Cheng Feng, Li Kong, Xudong Tian and Junsheng Zhou . 24211

Compute Allocation in Evolutionary Search: From Depth–Breadth to Multi-Armed Bandits
Sixue Xing, Haoyu He, Kerui Wu, Zhuo Yang, Haozheng Luo, Tianfan Fu and Aarthy Nagarajan 24224

Confidence Before Answering: A Paradigm Shift for Efficient LLM Uncertainty Estimation
Changcheng Li, Jiancan Wu, Hengheng Zhang, Zhengsu Chen, Guo An, Junxiang Qiu, Xiang Wang and Qi Tian 24252

A Comprehensive Survey on Linguistic Steganography: Methods, Countermeasures, Evaluation, and Challenges
Ruiyi Yan, Chenhui Chu, Zhongliang Yang and Yugo Murawaki 24271

Beyond Output-Level Correction: Tracing Positional Bias to Early-Layer Commitment in Transformer Multiple-Choice Evaluation
Soyeon Lee and Minhyeok Lee 24307

SocialArena: Evaluating Social Intelligence in Large Language Models
Zongqi Wang, Rui Wang, Yuchuan Wu, Hua Yang, Bingxue Qiu, Tianle Gu, Bing Zhao, Yujiu Yang, Jieping Ye and Yongbin Li 24329

Visual Compliance via Executable Safety Rule Entailment
Jisoo Kim, TaeYoon Kwack, Jinwoo Jang and Honguk Woo 24356

When Counterfactual Debiasing Backfires: Identity Cues as Task-Dependent Evidence in Harmful Language Understanding
Jia Duan, Xiaoyu Lu, Wenjing Hu and Zihao Ding 24392

KV Admission: Learning What to Write for Efficient Long-Context LLM Inference
Yen-Chieh Huang, Pi-Cheng Hsiu, Rui Fang and Ming-syan Chen 24411

Seeing is Believing? Evaluating Vision-Language Model Susceptibility in Agent-to-Agent Multimodal Persuasion
Haoyi Qiu, Yilun Zhou, Pranav Narayanan Venkit, Kung-Hsiang Huang, Jiaxin Zhang, Nanyun Peng and Chien-Sheng Wu 24440

How Stable Is In-Context Learning? Lipschitz Bounds for Transformer Output under Context Perturbation
Jinheng Wu and Di Zhang 24469

Beyond Outcome Gaps: Process-Aware Fairness Diagnosis for LLM-based Multi-Agent Decision Systems
Yiran Zhao, Lu Zhou, Liming Fang, Yufei Chen, Jiafei Wu, Zhe Liu and Xiaogang Xu 24481

RASPER: Reward-Aligned Summarization of Clinical Notes for EHR Outcome Prediction
Arya Hadizadeh Moghaddam, Mohsen Nayebi Kerdabadi, Chen Chen, Dongjie Wang and Zijun Yao 24519

LatCom: Cross-Agent Latent Compression for Efficient Multi-Agent Collaboration
Shinan Zhang, Tao Zhang, Qihui Zhu, Mengjie Zhang, Dong Jin, Yunpeng Hou, Shuangwu Chen, Xiaobin Tan, Quan Zheng and Jianyang 24531

Authority Bias in Conversational Search Engines for Academic Paper Recommendation.
Uthman Jinadu, Parsa Ghazvinian, Anjila Budathoki, Benjamin M. Ampel, Rajshekhar Sunderraman and Yi Ding 24548

RIPPLE: Attacking Backdoor-based LLM Fingerprinting via Latent Representation Perturbation
Hang fu, Yinghan Zhou, Wanli Peng, Juan Wen, Yiming Xue and Huimin Pei 24569

JPO: Juris Policy Optimization for Structured Legal Reasoning in Criminal Judgment Prediction
Zhaolu Kang, Yantao Liu, Tailong Luo, Leqi Zheng, Lei Wei, Chenghua Zhu, Junhao Gong, Jiachen Qian, Eric Hanchen Jiang, Jiaxin Liu, Yuan Wang, Hao Zhang, Zixia Wang, Rong Fu, Zheng Lin, Richeng Xuan and Zhichao Hu 24588

Mechanism Shift During Post-training from Autoregressive to Masked Diffusion Language Models
Injin Kong, Hyoungjoon Lee and Yohan Jo 24615

StepGuard: Learning Step-Level Guardrails with Scalable Supervision and Safety–Utility Balancing
Zhijie Zheng, Yu Li, Chen Qian, Yuqian Fu, Yanwei Fu, Lu Sheng, Jing Shao and Dongrui Liu 24644

ReaderScope: From Offline Reader-Response Calibration to Online Controlled Story Generation
Feng. Zhang and Rong Pan 24668

BiG-SURE - Bipartite Graph for Semantic Uncertainty and Reliability Estimation of LLMs
Debarpan Bhattacharya, Malay Phadke and Sriram Ganapathy 24713

WnW: Waxing-and-Waning KV Cache for Long-Form Speech LLMs
Yiming Yao, Chenyang Lyu, Xuanfan Ni, Longyue Wang, Weihua Luo, Yazheng Yang and Jinsong Su 24735

RedNote-Vibe: A Lifestyle Social Media Dataset for Detecting AI-Generated Text and Analyzing Its Impact
Yudong Li, Yufei Sun, Peiru Yang, Yuhan Yao, Wanyue Li, Haoyang Yang, Jiajun Zou, Haotian Gan, Linlin Shen and Yongfeng Huang 24749

Recognition–Refusal Misalignment in LLMs: Why Models Answer Structurally Unanswerable Questions
Yucheng Du and Xiyang Hu 24766

PredNav: A Predictive Cognitive Map for Zero-Shot Object Goal Navigation
Qingao Chai, Jiashuo Wang, Rui Yan and Huajin Tang 24795

ShapleyLaw: A Game-Theoretic Approach to Multilingual Scaling Laws
Xuyang Cao, Qianying Liu, Chuan Xiao, Yusuke Oda, Jiayi Wang, Pontus Stenetorp, Daisuke Kawahara, Makoto Onizuka, Sadao Kurohashi and Shuyuan Zheng 24812

Gendered Prompting and LLM Code Review: How Gender Cues in the Prompt Shape Code Quality and Evaluation
Lynn Janzen, Üveys Eroglu, Dorothea Kolossa, Pia Knoeferle, Sebastian Möller, Vera Schmitt and Veronika Solopova 24834

Concise and Sufficient Sub-Sentence Citations for Retrieval-Augmented Generation
Guo Chen, Qiuyuan Li, Qiuxian Li, Hongliang Dai, Xiang Chen and Piji Li 24853

Evading Fake News Detection: Self-evolving Social Manipulation via LLM-driven Agent Orchestration
Mengyang Chen, Lingwei Wei, Wei Zhou and Songlin Hu 24869

Meerkat2: Bridging Global Reasoning and Local Knowledge for Efficient Medical AI
Jungwoo Park, Hyeongsoon Hwang, Sanghoon Lee and Jaewoo Kang 24901

When Seekers Are Hard to Help: Evaluating Emotional Support Dialogue Systems in Worst-Case Interactions
Jiajie Yang, Yangchun Li, Guanyi Chen, Rui Fan, Xin Bai and Tingting He 24920

Deeper is Not Always Better: Mitigating the Alignment Tax via Confident Layer Decoding
Sining Zhoubian, Xuanming Zhang, Yuxuan Chen, Tianyi Tang, An Yang, Sean Du, Chujie Zheng, Fei Huang, Dayiheng Liu, Gao Huang and Jingren Zhou 24944

Beyond the Rabbit Hole: Mapping the Relational Harms of QAnon Radicalization
Bich Ngoc Doan, Giuseppe Russo and Gianmarco De Francisci Morales 24971

Beyond Text Following: Repairable Arbitration Reversals in Audio-Language Models
Yichen Gao, Yiqun Zhang, Zijing Wang, Yujia Li, Heng Guo, Xi Wu, Xiaocui Yang, Shi Feng, Yifei Zhang and Daling Wang 24992

SAILRec: Steering LLM Attention to Dual-Side Semantically Aligned Collaborative Embeddings for Recommendation
Xi Wu, Jiale Wang, Zihan Wang, Yichen Gao, Xiaocui Yang, Shi Feng, Daling Wang and Yifei Zhang 25013

OptiMer: Optimal Distribution Vector Merging Is Better than Data Mixing for Continual Pre-Training
Haiyue Song, Huy Hien Vu and Masao Utiyama 25034

Measuring Source-Resolution Distortion in LLM Output Arbitration: A Source-Bounded Evaluation Framework
Jinning Zhang, Yeqiang Wang, Yulong Zhang and Xuanbo Su 25055

Identify, Classify, Act: Per-Instruction Controlled Compliance for LLMs over External Data
Zesheng Shi, Xue li, Chenrui Dai and Jing Li 25072

Sampling Less, Knowing More: Contradiction-aware Adaptive Spectral and Entropic Uncertainty Quantification for LLMs
Jiaming Zhang, Heng-yang Lu, Xinan Chen, Junyuan Zheng, Yuncheng Jiang, Ying Zhu and Chenyou Fan 25092

RubricRM: Generative Reward Modeling via Dynamic Rubrics for Image Generation and Editing
Zijian Kan, Wei Wang, Luo Long, Bing Zhao, Xuan Ren, Weixu Qiao, Li Wenbo, hu Wei and Lin Qu 25107

HARVE: Hacking-Aware Reward-Head Vector Editing for Robust Reward Models
Shuang Liu, Yuxuan Bo, Qiuyang Zhao, Caiyue Huang, Xiaorong Chen, Yanguang Liu and Mengnan Du 25127

From Content Generation to Learning Support: Pedagogy-Guided Generative Video Tutors for STEM Learning
Xinchen Ma, Shuimu Wang, Gaole He, Yanbin Zhang, Chunyang Wang, Yunshi Lan and Weining Qian 25151

MAViQ: Memory-Augmented Visual Querying for Multi-Turn MLLM Jailbreaks
JinYun Lu, Xiaodan Li, Guanghao Zhou, Cen Chen and YueFeng Chen 25170

CoreMem: Riemannian Retrieval and Fisher-Guided Distillation for Long-Term Memory in Dialogue Agents
Jiaqi Chen, Yongqin Zeng, Shaoxiong Zhan, Shaoshen Chen, Yijian Zhang, Chunxia Ma, XiuTeng Zhou and Hai-Tao Zheng 25189

Small Models, Fewer Heads, More Forgetting: Sparse Attention Fine-Tuning for SLM Continual Learning
Hao Xu, Shezheng Song, Long Peng, Xiaodong Liu, Xiaopeng Li, Ma Jun, Shasha Li and Jie Yu 25205

Lung-R1: A Knowledge Graph-Guided LLM for Pulmonary Diagnostic Reasoning
Haoyang Zeng, Yuanxi Fu, Rongzhen Li, Ymyang, Xiao Sun, Jingwang Huang, Gujie Shao, Guohui Xiang, Quan Lu, Ye Dongfan, Xuetao Chen, Jiang Zhong, Xu Zhi and Kaiwen Wei 25223

Rigel: Self-Distilled Score Adaptation for Image and Video Captioning Evaluation
Shuitsu Koyama, Kazuki Matsuda, Yuiga Wada, Shinnosuke Hirano, Daichi Yashima and Komei Sugiura 25245

SecOPD: Mitigating Adaptive Prompt Injections by On-Policy Distillation
Yibo Peng, Long Lian, David Wagner and Sizhe Chen 25262

ObSchema: From Local Observations to Global Entity Attribute Schemas in Scientific Literature
Xueqiang Xu, Wonbin Kweon, Jimeng Shi, Runchu Tian, Yunfan Kang, Wei Hu, Shaowen Wang and Jiawei Han 25278

HLoRA: Hierarchical Regularization on Combined Parameter Space for LoRA Fine-Tuning
Shezheng Song, Hao Xu, Long Peng, Xiaopeng Li, Shasha Li and Jie Yu 25301

APEx: Distillation of Agent Procedural Experience for Adaptive Deep Research Question Answering
Jie Ding, Rui Sun, Xinyuan Zhang, Zeyu Zhang and Xin Liu 25317

Beyond Outliers: A Data-Free Layer-wise Mixed-Precision Quantization Approach Driven by Numerical and Structural Dual-Sensitivity
Hengyuan Zhang, Xinrong Chen, Zunhai Su, Xiao Liang, Jing Xiong, Wendong XU, He Xiao, Chaofan Tao, Wei Zhang, Ruobing Xie, Lei Jiang, Hayden Kwok-Hay So and Ngai Wong 25342

AOR-Bench: Do Large Audio Language Models Over-Refuse Pseudo-Harmful Queries?
Jiaxi Yang, Chaewan Chun, Jason S Lucas, Yuchen Yang and Dongwon Lee 25360

LexPath: A Domain-Oriented Multi-Path Framework for Legal Article Retrieval
Weixuan Liu, Qingfeng Zhuge and Xuyang Chen 25377

Fathom-Vaidya: Advancing Medical Reasoning with Rubric-Based Rewards
Kalash Shah, Kunal Singh, J Snehan and Shreyas Singh 25395

ANI: Adaptive Numerical Injection for Unifying Semantic and Arithmetic Representations in Numerical Reasoning
Jinsung Jeon and Seung-Won Hwang 25414

A Unified Mechanistic Analysis of Knowledge- and Safety-Based Refusals
Yuri Son, Seunghee Kim, Hyuhng Joon Kim and Taeuk Kim 25430

The Straight and Narrow: Do LLMs Possess an Internal Moral Path?
Luoming Hu, Jingjie Zeng, Liang Yang and Hongfei Lin 25460

Can LLMs Reliably Self-Report Adversarial Prefills, and How?
Quang Minh Nguyen, Uzair Ahmed and Taegyoon Kim 25483

Understanding the Effects of Distractors on Reasoning Vision-Language Models
Jiyun Bae, Hyunjong Ok, Sangwoo Mo and Jaeho Lee 25513

Rubric-Aware Multi-Expert Online Policy Distillation for Chinese Elementary School Essay Writing
Wei Tian, Yuhao Zhou, Xiayu Sun, Yadong Zhang, Hao Zhou, Shu Liu and Man Lan 25543

FlowMM: Cross-Modal Information Flow Guided KV Cache Merging for Efficient Multimodal Context Inference
Kunxi Li, Yufan Xiong, Zhonghua Jiang, Yiyun Zhou, Zhaode Wang, Chengfei lv and Shengyu Zhang 25566

DeepSearch-World: Self-Distillation for Deep Search Agents in a Verifiable Environment
Xinyu Geng, Xuanhua He, Sixiang Chen, Yanjing Xiao, Fan Zhang, Shijue Huang, Haitao Mi, Zhenwen Liang, Tianqing Fang and Yi R. Fung 25576

OpenSkill: Open-World Self-Evolution for LLM Agents
Zhiling Yan, Dingjie Song, Hanrong Zhang, Wei Liang, Yuxuan Zhang, Yutong Dai, Lifang He, Philip S. Yu, Ran Xu, Xiang Li and Lichao Sun 25602

ERR+: Sequential Entropy Resolution for Efficient and Decisive LLM Reasoning
Xin Jiang, Minhao Wang, Wen Wu, Zhentao Xie, Shangheng Du, Jinxin Shi, Jiabao Zhao, Liang He and Weicong Chen 25619

Touch What You Say: Contact-Local Tactile Grounding for Physical-Property Instructions
Le Wang, Kelan Peng, Zhang Chenyu and Wenjuan Han 25635

InterPol: De-anonymizing LM Arena via Interpolated Preference Learning
Minsung Cho and Jaehyung Kim 25649

Training-free LLM Verification via Recycling Few-shot Examples
Dongseok Lee, Jimyung Hong, Dongyoung Kim and Jaehyung Kim 25669

What Drives Paper Acceptance? A Process-Centric Analysis of Modern Peer Review
Inbum Heo, Goun Pyeon, Hyungjin Ahn and Sangkeun Jung 25711

Not What, But How: A Communicative Audit of LLM Response Framing
Siddhesh Milind Pawar, Sarah Masud, Haneul Yoo, Alice Oh and Isabelle Augenstein 25732

Jasper-Flash: Elastic Token Compression and Progressive Distillation for Inference-Scalable Text Embedding Models
Haining An, Xingbing Chen, Wenbo Lyu, Dun Zhang, Ziyang Zeng, Zhenxiang Zhao, Rui Song and Hao Xu 25766

Speaking While Listening: A Survey and Empirical Audit of Full-Duplex Spoken Dialogue Systems
Jingyu Lu, Yuhan Wang, Jianming Luo, Yifu Chen, Tianle Liang, Shengpeng Ji, Ziyue Jiang, Xiaoda Yang, Yu Zhang, Xize Cheng, Chenyuhao Wen, Changhao Pan, Haoxiao Wang, Chen Ye, Jian Wu, Xiaoxi Jiang, Guanjun Jiang and Zhou Zhao 25790

PAC: Progress-Augmented Advantage Curriculum for Multi-Task Reinforcement Learning of LLMs
Yuanqiang Yu, Yanzhao Zheng, Zhentao Zhang, Tianze Xu, Chao Ma, Jihuai Zhu, Jiashun Liu, Xinle Deng, Baohua Dong, Hangcheng Zhu and Ruohui Huang 25825

Latent Thoughts, Certain Gains: Bayesian Optimization for LLM Reasoning
Xin Xiao and Jiang Zhong 25844

Can MLLMs Perceive and Describe Medical Image Quality Beyond Scores?
Jiyao Liu, Jinjie Wei, Wanying Qu, Chenglong Ma, Junzhi Ning, Yunheng Li, Ying Chen, Jiafan Zhang, Pengcheng Chen, Xin Gao, Ming Hu, Xin Wang, Shujian Gao, Dingkang Yang, Zhongying Deng, Jin Ye, Lihao Liu, Junjun He and Ningsheng Xu 25854

Branch Later, Compute Less, Think Better: Adaptive Depth Parallelism for Efficient Parallel Prefix-Tuning
Mingqian Zhang, Mouxiang Chen and Zhongxin Liu 25878

SCOPE: Signal-Calibrated On-Policy Distillation Enhancement with Dual-Path Adaptive Weighting
Binbin Zheng, Xing Ma, Yiheng Liang, Jingqing Ruan, Xiaoliang Fu, Kepeng Lin, Benchang Zhu, Ke Zeng and Xunliang Cai 25897

UltraX: Refining Pre-Training Data at Scale with Adaptive Programmatic Editing
Xinlong Zhao, Dongsheng Liu, Hengyu Zhao, Zixuan Fu, Zheng Wang, Jie Cai, Jie Zhou, Qiang Ma, Xuanhe Zhou, Xu Han, Yudong Wang and Zhiyuan Liu 25915

PsychēChat: An Empathic Framework Focused on Emotion Shift Tracking and Safety Risk Analysis in Psychological Counseling
Zhentao Xia, Yongqi Fan, Yuxiang Chu, Yichao Yin, Liangliang Chen, Tong Ruan and Weiyan Zhang 25953

SpecAlign: Efficient Specification-Grounded Alignment of Large Language Models via Synthetic Data
Wenjie Wang, Yue Huang, Zhengqing Yuan, Han Bao, Shiyi Du, Yuchen Ma, Yue Zhao, Yanfang Ye and Xiangliang Zhang 25986

SimulS2ST-Omni: Data-Efficient Streaming Speech-to-Speech Translation via Explicit Trajectory Supervision
Rongshen He, Xinyu Liang, Dekun Chen, Jiaqi Li, Mingjie Chen and Zhizheng Wu 26030

InFi-Check: Interpretable and Fine-Grained Fact-Checking of LLMs on Grounding Documents
Yuzhuo Bai, Shuzheng Si, Kangyang Luo, Qingyi Wang, Wenhao Li, Gang Chen, Fanchao Qi and Maosong Sun 26059

TRACER: An Information Entropy-Grounded Framework for Efficient Large-table Semantic Annotation
Miaoqin Zheng and Shuangyin Li 26088

MoARa: Module-Aware Rank Allocation and Structure-Preserving Decomposition for Low-Rank LLM Pre-training
Keunyoung Kim and Nojun Kwak 26098

Draft-OPD: On-Policy Distillation for Speculative Draft Models
Haodi Lei, Yafu Li, Haoran Zhang, Shunkai Zhang, Qianjia Cheng, Xiaoye Qu, Ganqu Cui, Bowen Zhou, Ning Ding, Yun Luo and Yu Cheng 26128

CaptchaMind: Training CAPTCHA Solvers via Reinforcement Learning with Explicit Reasoning Supervision
Pengcheng Wang, Haoxiang Liu, Yang Dai, Xiangxiang Zeng, Guanhua Chen, Baotian Hu, Longyue Wang and Weihua Luo 26142

SecAgent: Efficient Mobile GUI Agent with Semantic Context
Yiping Xie, Chensong, Jingxuan Xing, Wei Jiang, Zekun Zhu, Yingyao Wang, Pi Bu, Jun Song, Yuning Jiang and Bo Zheng 26159

VISA: Video Skeleton-Aware Efficient Frame Selection for Long Video Understanding
Qihui Zhu, Zijian Wen, Tao Zhang, Mengjie Zhang, Shenghao Ye, Dong Jin, Shuangwu Chen, Xiaobin Tan, Quan Zheng and Jianyang 26175

Polished but Unresolved: Identifying Late-Stage Pressure States in Long-Horizon Tool-Use Agents
Haoyang Chen, Yi Liu, Jianzhi Shao, Xiaozhou Xu, Zhe Sun and Wei Hu 26189

When Calibration Rankings Reverse: Accuracy-Controlled Evaluation for Fair Comparison of LLMs
Zhichao Yang, Caiqi Zhang, Ruihan Yang, Chengzu Li, Nigel Collier and Deqing Yang . . . 26206

One Model, Many Graphs: Learning over Attributed Graphs across Heterogeneous Modalities with Vision-Language Models
Jiayi Yang, Yifang Chen, Yuanfu Sun, Jiajin Liu and Qiaoyu Tan 26232

CAVD-Guard: Defending Against Token Fragmentation Jailbreaks via Attention Footprints
Junjie Xu and Zicheng Zhao 26254

Reinforcement Learning Improves Traversal of Parametric Knowledge in LLMs
Renfei Zhang, Manasa Kaniselvan, Rylan Schaeffer and Niloofar Mireshghallah 26282

Voices Across Registers: Corpus-Conditioned Vernacular Jailbreaks against Aligned LLMs via Fanfiction Subgenres
Zhongze Luo, Ruihe Shi, Zhenshuai Yin, Haoyue Liu, Weixuan Wan and Xiaoying Tang . . 26314

Patient-Zero: Scaling Synthetic Patient Agents to Real-World Distributions without Real Patient Data
Yunghwei Lai, Ziyue Wang, Weizhi Ma and Yang Liu 26339

Distribution-aware Language Neuron Identification in Multilingual Large Language Models
Minjun Kim, Inho Won, Junghun Yuk, Dongyeon Kim, Jihyo Kim and KyungTae Lim 26366

SpeechEditBench: A Bilingual Multi-Attribute Benchmark for Instruction-Guided Speech Editing
Hanlin Zhang, Daxin Tan, Dehua Tao, Xiao Chen, Haochen Tan and Linqi Song 26394

MedF^{2}Route: Evidence-Routed Graph Verification for Hallucination Detection in Clinical LVLMs
Xiao Song, Ruiying Chen, Haonan Qin, Jianyuan Huang, Yuqi Fang and Caifeng Shan . . . 26415

Don't Read Everything: A Curvature-Conditioned Query for Linear Attention
Dong Le, Thong Thanh Nguyen, Cong-Duy T Nguyen and Anh Tuan Luu 26439

LongAct: Harnessing Intrinsic Activation Patterns for Long-Context Reinforcement Learning
Bowen Ping, Zijun Chen, Tingfeng Hui, Qize Yu, Chenxuan Li, Junchi Yan and Baobao Chang 26459

Visual Attention Faithfulness in Vision-Language Models is Heterogeneous
Xurui Song, Weishi Wang, Zhongqi Yue, Kuluhan Binici, Tao Bai, Hongxin Shao, Daniel Dahlmeier and Jun Luo 26472

Stable Reasoning, Unstable Responses: Mitigating LLM Deception via Stability Asymmetry
Guoxi Zhang, Jiawei Chen, Tianzhuo Yang, Lang Qin, Juntao Dai, Yaodong Yang and Jingwei Yi 26487

MedPriv-Bench: Benchmarking the Privacy-Utility Trade-off of Large Language Models in Medical Open-Ended Question Answering
Shaowei Guan, Zhai Yu, Hin-Chi Kwok, Jiawei Du, Xinyu Feng, Jing Li, Jing Qin and Vivian Hui 26512

EvoGenUI-Bench: Evaluating LLMs as Multi-Turn Generative UI Assistants
Yue Peng, Lanke Xia, Zihan Wang, Jiahao Ye, Ke Ning and Hongyi Wen 26533

Cognitive Chunking for Soft Prompts: Accelerating Compressor Learning via Block-wise Causal Masking
Guojie Liu, Yiqi Wang, Yanfeng Yang, Wenqi Fan, Songlei Jian, Jianfeng Zhang and Jie Yu26550

Select, Label, Evaluate: Active Testing in NLP
Antonio Purificato, Maria Sofia Bucarelli, Andrea Bacciu, Fabrizio Silvestri and Amin Mantrach 26566

Correlation-Flip Visual Injection for Hallucination Mitigation in Large Vision-Language Models
Jiaxuan Cao, Zeyu Shi, Longhua Li, Lei Wang and Lei Qi 26585

The Geometry of Safety Drifts: Step-level Refusal-Subspace Monitoring for Reasoning Large Language Models
Liming Lu, Anan Du, Fanzhen Liu, Hechang Chen and Shuchao Pang 26597

Streaming Atomic Action Segmentation in Egocentric Videos via Iterative LVLM-to-Streaming Distillation
Hyolim Kang, Jungsoo Lee, Janghoon Cho, Juntae Lee, Seon Joo Kim and Sungha Choi . . 26613

CROG: Coverage-aware Reasoning Optimization on Graphs for Multi-hop Question Answering
Junli Liang, Pengfei Zhou, Rui Lv, Wangqiu Zhou, Qi Song and Xiang-Yang Li 26632

MMDS-Bench: Benchmarking Multimodal Large Language Models on Dynamic Stance in Social Media Interactions
Yuzhe Ding, Kang He, Li Zheng, Shengwu Zheng, Teng Shi, Fei Li, Chong Teng and Donghong Ji 26646

MMGR: Multi-Modal Generative Reasoning Benchmark and Evaluation
Zefan Cai, Haoyi Qiu, Tianyi Ma, Haozhe Zhao, Gengze Zhou, Tingting Liao, Xinyan Velocity Yu, Kung-Hsiang Huang, Ke Wan, Shawn Lin, Parisa Kordjamshidi, Minjia Zhang, Wen Xiao, Jiuxiang Gu, Nanyun Peng and Junjie Hu 26668

EmoDap: Dual-Agent Prompt Retrieval Framework for Zero-Shot Fine-Grained Emotion Recognition
Kun Peng, Cong Cao, Hao Peng, Xiaoyan Yu, Zihao Wang, Lei Jiang, Yu Liu, Yanbing Liu and Philip S. Yu 26752

OP-Bench: Benchmarking Over-Personalization for Memory-Augmented Personalized Conversational Agents
Yulin Hu, Zimo Long, Jiahe Guo, Xingyu Sui, Xing Fu, Weixiang Zhao, Yanyan Zhao and Bing Qin 26770

MM-BrowseComp: A Comprehensive Benchmark for Multimodal Browsing Agents
Shilong Li, Xingyuan Bu, Wenjie Wang, Jiaheng Liu, Jun Dong, Haoyang He, Hao Lu, Haozhe Zhang, Chenchen Jing, Zhen Li, Chuanhao Li, Jiayi Tian, Chenchen Zhang, Tianhao Peng, Yancheng

He, Jihao Gu, Hui Huang, Donghao Zhou, Yuanxing Zhang, Jian Yang, Ge Zhang, Wenhao Huang, Zhaoxiang Zhang, Qiangpeng Yang and Shilei Wen 26828

RAVE: Re-Allocating Visual Attention in Large Multimodal Models
XI Leng, Xinhong Ma, Ziqiang Dong, Feng Zhang, Xiaoying Tang, Yang Yang and Guanjun Jiang 26848

Whether LLMs Can Navigate Beliefs and Facts Depends on How You Phrase It
Quang Minh Nguyen and Luis Frentzen Salim 26862

QUORUM: QUality-Optimized Routing Using Multiple annotators
Antonio Purificato, Maria Sofia Bucarelli, Andrea Bacciu, Amin Mantrach and Fabrizio Silvestri 26873

Bottom-up Policy Optimization: Your Language Model Policy Secretly Contains Internal Policies
Yuqiao Tan, Minzheng Wang, Shizhu He, Huanxuan Liao, Chengfeng Zhao, Tian Liang, Jun Zhao and Kang Liu 26891

TRACE: An Evidence-Grounded Benchmark for Safety Evaluation of Large Reasoning Models
Zhenyu Wu, Siyuan Chen, Changchun Yang, Jiaqi Dong, Min Zhou, Ali A Almadan, Talal Alhammad, Faisal Wahbo, Aminullah Tora, Mona Alshahrani and Xin Gao 26911

CRAM: Centroid-Routing and Adaptive MoE for Multimodal Continual Instruction Tuning
Jun-Tao Tang, Zhen-Hao Xie, Yu-Cheng Shi and Da-Wei Zhou 26936

Risky Business: Measuring The Faithfulness-Safety Tension
Dominik Meier, Luca Joshua Francis, Marco Bernhard Kaiser, Terry Ruas, Jan Philip Wahle and Bela Gipp 26954

SkillPyramid: A Hierarchical Skill Consolidation Framework for Self-Evolving Agents
Yuan Xiong, Ziqi Miao, Chen Qian, Lijun Li, Yequan Wang, Shizhu He, Jun Zhao and Kang Liu 26974

SARL-ICD: Structure-aware Contrastive Representation Learning with LLM-augmented Label Descriptions for Automatic ICD Coding
Jincheng Shi, Tianyang Huang and Shanfeng Zhu 26993

MetaBench: Benchmarking Two-Stage Reranking and Top-N Screening for Meta-Analyses
Yining Huang, Keke Tang and Meilian Chen 27010

PRISM: Entropy-Guided Multi-View Exploration for Scientific Idea Generation
Jiajun Yu, Weifan Mao, Kaipeng Zeng, Kejia Bian, Zhuo Wang, Kun Li and Haishuai Wang27051

Revealing Multi-View Hallucination in Large Vision-Language Models
Wooje Park, Insu Lee, Soohyun Kim, Jaeyun Jang, Minyoung Noh, Kyuhong Shim and Byonghyo Shim 27073

SHIFT: Self-reconstruction Harnesses Implicit Fine-grained Thinking for Retrieval
Yuxiao Luo, Da Li, Mingjie Zhang, Zhentao he, Shikun Zhang and Wei Ye 27103

LKV: End-to-End Learning of Head-wise Budgets and Token Selection for LLM KV Cache Eviction
Enshuai Zhou, Yifan Hao, Chao Wang, Rui Zhang, Di Huang, Jiaming Guo, Xing Hu, Zidong Du, Qi Guo and Yunji Chen 27116

Affective Context Amplifies Sycophancy in LLM Responses
Jiayi Li, Sanjana Menon, Brett Frischmann, Shomir Wilson and Sarah Rajtmajer 27134

HiTeC: Hierarchical Contrastive Learning on Text-Attributed Hypergraph with Semantic-Aware Augmentation
Mengting Pan, Fan Li, Chen Chen, Xiaoyang Wang and Wenjie Zhang 27157

CYPHERTURN: A Multi-Turn Benchmark for Conversational Text-to-Cypher Evaluation and the Autonomy Divergence
Yuzhe Zhang, Weijie Zhu, Haolin Yang, Ziyun Zhang, Xianwei Xue, Mengke Chen, Qiutong Pan and Huaqian Cai 27173

MLLMCLIP: Feature-Level Distillation of MLLM for Robust Vision-Language Representations
Jongsuk Kim, Qiyu Wu, Zhuoyuan Mao, Hiromi Wakaki, Junmo Kim and Yuki Mitsufuji . 27208

CoDe-LoRA: Mitigating the Orthogonality Dilemma in Continual Learning of LLMs via Knowledge Consolidation and Decoupling
Maoqi Liu, Quan Fang and Yufei He 27223

Supervised Fine-Tuning for Reasoning Large Language Models: A Survey
Yongcan Yu, Kuangpu Guo, Jian Liang, Lingxiao He, Meng Wang and Ran He 27240

LADA: Enabling Adaptation of Black-Box LLMs to Dynamic Domain Changes at Test Time
Yihao Hu, Congyu Qiao, Junming Yang, Mofei Song, Xin Geng and Ning Xu 27274

Rethinking the Role of Temperature in Large Language Model Distillation
Hoang-Chau Luong and Lingwei Chen 27292

ACTD: Anchor-Based Cross-Tokenizer Distillation with Residual Regularization
Huiyi Zhang, Zijian Li, Xiaocheng Feng, Weitao Ma, Xiaoliang Yang, Yichong Huang and Bing Qin 27304

Reflect to Inform: Boosting Multimodal Reasoning via Information-Gain-Driven Verification
Shuai Lv, Chang Liu, Feng Tang, Rui Chen, Yu-Jie Yuan, Aojun Zhou, Kui Zhang, Xi Yang and Yangqiu Song 27318

COLLABMEM: A Diagnostic Benchmark for Collaborative Memory Operations in Multi-Agent LLM Systems
Shun Wu, Di Wu, XueYou Zhang and Kang Liu 27343

Over-reasoning as a Vulnerability: Neuron Activation-Guided Denial-of-Service Attacks on Large Reasoning Models
Jiaxiang Li, Zhipeng Wei, Tianwen Qian, Yuanbin Wu, Xianming Hu and Xiaoling Wang . 27357

VisCache: Visual KV Cache Pruning for Efficient Vision Large Language Model Inference
Lyuke Wang, Zhuo Li and Guangxu Zhu 27385

Retrievable Gradients: Continual Post-Training Without Cumulative Weight Drift
Weihang Su, Jiacheng Kang, Jingyan Xu, Qingyao Ai, Jianming Long, Hanwen Zhang, Bangde Du, Xinyuan Cao, Min Zhang and Yiqun Liu 27407

Imagine-then-Plan: Agent Learning from Adaptive Lookahead with World Models
Youwei Liu, Jian Wang, Hanlin Wang, Beichen Guo and Wenjie Li 27427

Does Alignment Debias or Just Suppress? Evaluating Stereotypical Bias Across Base-Instruct Pairs
Sigurd Kornelius Havstein and Anders Søgaard 27446

CoMAP: Co-Evolving World Models and Agent Policies for LLM Agents
Youwei Liu, Jian Wang, Hanlin Wang and Wenjie Li 27461

Reading Time Modeling with Pixel-Based Language Models
Junlin Li, Sara Alhajeri and Tatsuki Kuribayashi 27481

Controlling Creativity in Concept Space: A Framework for Generative Models
Dana Levin and Dafna Shahaf 27501

Linguistics-Aware Non-Distortionary LLM Watermarking
Shinwoo Park, Hyejin Park, Hyeseon An and Yo-Sub Han 27521

REPAIR: Resolving Long-Tail Confusion in Scientific Retrievers via Fact-Verified Iterative Refinement
Yerim Oh and Gunhee Kim 27543

EMER-Guard: Protocol-Grounded Verification for LLM Emergency Instructions
Zida Yang, Guquan You and Yu Zheng 27573

Complexity-Aware Routing: Detecting and Allocating Dynamic Compute for Efficient Mixture-of-Experts Large Language Models
Jiahe Li, Saisai Hu, Menghao Li, Shaolin Zhu and Deyi Xiong 27594

Advancing Fact-Checking for Large Language Models via Argumentative Reasoning
Zirong Chen, Xiangwei Lv, Zhaoqun Li, Wenwei Ke, Jinxiong Chang, Linjian Mo and Beishui Liao 27611

DCGC: Draft-Conditioned Global Correction for Complex Reasoning with Masked Diffusion Models
Minhae Oh, Nakyung Lee and Jungwoo Lee 27631

Large-scale Multilingual News Image Captioning with LLMs
Yuji Chen, Purushoth Velayuthan, Alistair Plum, Hansi Hettiarachchi, Saroj Basnet, Menan Velayuthan, Marcos Zampieri and Tharindu Ranasinghe 27652

MDHB: A Multi-Dimensional Online Harm Detection Benchmark for Proactive Psychosocial Risk Forecasting
Yuting Huang, Xiaoyong Huang, Hao Wang, Heli Sun, Jiaruo Wu, Zejun Wang, Yunyun Shi, Guanlan Yue and Liang He 27663

Where to Look and What to Use: Retrieve–Localize–Generate for Long-Term Conversational Memory Question Answering
Yifan Wang, Xinkui Lin, Yongxiu Xu, Shen Gao, Ruochen Yang, Kun Huang, Yubin Wang, Jie Wu, Wei Liu, Jian Luan, Hongbo Xu and Shuo Shang 27680

Reservoir of Importance: Learning Semi-Structured Sparsity with Differentiable Subset Sampling
Ha Dinh, Xuan Duy Ta, Khoat Than and Khac-Hoai Nam Bui 27702

Criterion-Activated Preference Synthesis: Attribution-Preserving Preference Data from Explicit Standards
Zian Wang, Pu Lu, Caishuang Huang, Wuwenxi, Meng Zhou, Xiku Du and Yang Qiao . . . 27719

Teaching by Learning: Harnessing LLMs Collaboration for Adaptive Data Synthesis via Cognitive Load Feedback
Yue Luo, Feng Zhao, Yu Yang and Guandong Xu 27739

ANCHORS: Adaptive Normalization for Continual High-Order Retention via Sigmoid-Guided Magnitude Control
Yuxuan Zhang and Jiaqi Zhu 27755

TimeSense: Decoupling Modality Learning for Robust Time Series Reasoning
Zhirui Zhang, Changhua Pei, Tianyi. Gao, Zhe Xie, Yibo Hao, Zhaoyang Yu, Longlong Xu, Tong Xiao, Jing Han, Gaogang Xie and Dan Pei 27772

CoRRSafe: Improving LRM Safety via Co-optimized Fine-Grained Reasoning and Response Alignment
Yangxiao Song, Bing Xu, Bufan Xu, Muyun Yang, Conghui Zhu, Huanqin Wu, Bojie Hu, Hongjiao Guan and Wenpeng Lu 27790

Rethinking Cross-Modality Prompt Transfer: Geometric Priors over Linguistic Semantics
Ningyuan Zhang, Jie Lu, Keqiuyin Li and Guangquan Zhang 27815

TCC-Bench: Benchmarking the Traditional Chinese Culture Understanding Capabilities of MLLMs
Pengju Xu, Yan Wang, Shuyuan Zhang, Xuan Zhou, Lixin, Yuanyue, Fengzhao Li, Shunyuan Zhou, Xingyu Wang, Yi Zhang and Haiying Zhao 27834

Language Chain in Alignment: Cross-lingual Ranking Preference Optimization
Seungyoon Lee, Minhyuk Kim, Jungseob Lee and Heuiseok Lim 27855

Q-CAT: A Non-Parametric Query Filtering System for Prohibited Content in Dynamic Enterprise Environments
Juheon Kang, Eunkyung Choi, Hongseok Oh, Hyunseo Shin, JuYoun Son, Yunjin Bae and Wonseok Hwang 27870

Rubric-Guided Process Reward for Stepwise Model Routing
Shenghao Ye, Yu Guo, Zhengheng Li, Qiang Huang, Shuangwu Chen, Wang Luo, Rui Qian, Jie Zhang, Tao Zhang, Dong Jin, Xiaobin Tan, Quan Zheng and Jianyang 27892

CacheSpec: Finding the Sweet Spot for Small Models in Large Language Models
Jingquan Chen, Jie Feng, Jinghua Piao, Shaogang Hu and Yong Li 27911

ScopeBench: Evaluating Cultural Norm Scope Awareness in LLMs
Haojun Liang, Ziwei Zhao, Chenyu Li, Qiyue Zhao, Rongduomijia Mu, Yulan Shen, Yu Gu, Chen Zhang, Kairui Zhang, Yanbiao Ma and Wenxuan Wang 27929

A Scaled-Up Empirical Study of Syntactic Language Models
Pengcheng Wang, Zhexuan Zhang, Chao Lou and Kewei Tu 27961

Shallow to Deep: Aligning Token Pruning with Stage-wise Roles in LVLMs
Shuo Zhang, Jintao Tong, Yixiong Zou, Yuhua Li and Ruixuan Li 27977

Towards a Comprehensive Scaling Law of Mixture-of-Experts
Guoliang Zhao, Yuhan Fu, Shuaipeng Li, Samm Sun, Ruobing Xie, An Wang, Weidong Han, Zhen Yang, Weixuan Sun, Yudong Zhang, Cheng-zhong Xu, Di Wang and Jie Jiang 27994

CulturePRM: A Process Reward Model for Mitigating Cultural Overriding in Cultural Reasoning
Ji-Eun Han and Yoonseok Heo 28023

Diagnose, Then Refine: A Closed-Loop TTS System with AudioLLM-guided Correction
Zeyang Song, Tianchi Liu, Tianrui Wang, Chenglin Xu, Steven Y. Guo and Haizhou Li . . . 28041

Hidden Threat in Synthetic Data: Covert Targeted Bias Injection through Benign Text
Minkyung Cho, Jihyo Kim, SeungWoo Song, Junghun Yuk, Minjoon Kee, Hoyun Song and KyungTae Lim 28060

Popular but Wrong: Understanding and Mitigating LLM Overconfidence through Knowledge Popularity
Shiyu Ni, Keping Bi, Jiafeng Guo and Xueqi Cheng 28085

Rethinking Stepwise Model Routing: A Cost-Efficient Table Reasoning Perspective
Shenghao Ye, Yuxiang Wang, Yu Guo, Tao Zhang, Shuangwu Chen, Zijian Wen, Qihui Zhu, Mengjie Zhang, Dong Jin, Wang Luo and Jianyang 28102

TraceProbe: An Execution-Based Benchmark for Diagnosing Safety Failures in LLM Agents
Yuxiang He, Jian Zhao, Yufei Liu, Yuchen Yuan and Xuelong Li 28121

Muon^2: Boosting Muon via Adaptive Second-Moment Preconditioning
Ziyue Liu, Ruijie Zhang, Zhengyang Wang, Yequan Zhao, Yupeng Su, Zi Yang and Zheng Zhang 28141

CAVEWOMAN: How Large Language Models Behave Under Linguistic Input and Output Compression
Morayo Danielle Adeyemi, Ryan A. Rossi and Franck Dernoncourt 28162

Why Act Without Asking? Evaluating Mobile GUI Agents under Ambiguous Instructions
Yeongjae Cho, Jungwuk Park, Tae-Min Choi, Chiyoun Park, Mincheol Kim and Kookhoi Kim 28189

AGORA: An Archive-Grounded Benchmark for Agentic Workplace Document Reasoning
Honglin Guo, Qi Zhang, Yu Zhang, Weijie Li, Rui Zheng, Zhikai Lei, Qiyuan Peng, Zhiheng Xi, Tao Gui and Qi Zhang 28211

A retrieval conditioned rebinding circuit for dynamic entity tracking in large language models
Soyoung Oh and Vera Demberg 28229

Syntactic Complexity Convergence in Dialogue: Analysis of the Phenomenon and Application to LLM Detection
Yu Wang, Yanran Chen, Yifan Wang, Steffen Eger and Hendrik Buschmeier 28244

When In-Distribution Gains Fail: Evaluating Weak-to-Strong Reward Models under Preference Shift
Khoi M. Le, Tri Cao, Phong Nam Huu Nguyen, Cong-Duy T Nguyen, Anh Tuan Luu, Chunyan Miao, See-Kiong Ng and Thong Thanh Nguyen 28270

SSR: Symbolic-Neural Soft Reasoning for Robust and Verifiable Thinking Chains
Rui Wang, Zeming Wei, Yihao Zhang and Xiaokun Luan 28289

LOKI: Locating Optimal Knock-In Points for Backdoor Attacks on Dense Retrievers
Qiqi Bao, Linzhi Feng, Yaguan Qian, Yanru Guo, Bin Wang, Zhaoquan Gu, Shouling Ji and Zhen Lei 28309

CODA: Mitigating Over-Refusal and Jailbreaks in Large Language Models via Codebook-Orthogonal Direction Alignment
Feng Huang, Yunqi Zhang, Qiang Huang, Qing Yu and Yongfeng Huang 28326

FOCAL: Fine-Grained Optimal-Transport-Driven Contrastive Alignment of Language and ECGs with Waveform Enhancement
Haitao Li, Che Liu, Zhengyao Ding, Ziyi Liu, Wenqi Shao and Zhengxing Huang 28349

VIRA-Bench: Diagnosing Pedagogical Contingency Understanding in Online Tutoring Videos
Weihang Wang, Hansu Gu, Guanming Liu, Peng Zhang, Ning Gu and Tun Lu 28368

A Family Affair: Counterintuitive transfer effects in ASR for endangered languages
Emily Prud'hommeaux, Hunter Scheppat, Nicholas Hernandez and Eric Le Ferrand 28389

CPR for LLMs: Critical-Point Routing against Catastrophic Forgetting in Domain Adaptation
Kwangmin Ki, Yunhun Nam, Jongheon Jeong and Jaehyung Kim 28398

X-Coder: Advancing Competitive Programming with Fully Synthetic Tasks, Solutions, and Tests
Jie Wu, Haoling Li, Xin Zhang, Jiani Guo, Jane Luo, Xuewei Yang, Steven Liu, Yangyu Huang, Ruihang Chu, Scarlett Li and Yujiu Yang 28423

AnyAudio-Judge: A Dynamic Rubric-Based Benchmark and Evaluator for Audio Instruction Following
Haitao Li, Tian Tan, Yang Yuguang, Shan Yang and Xie Chen 28451

ToolForge: A Data Synthesis Pipeline for Multi-Hop, Multi-Turn, and Self-Reflective Tool-Use Data
Hao Chen, Shunyuan Zhou, Zhexin Hu, Jiajun Chai, Hang He, Haocheng Yang, Xiaohan Wang, Wei Lin, Tianyu Wang, Zhuofeng Zhao and Guojun Yin 28470

IdioSteer: Controlling Figurative-vs-Literal Interpretation via Residual-Stream Activation Steering
Bar Cohen, Kai Golan Hashiloni and Kfir Bar 28493

Stop Before You Fail: Operational Capability Boundaries for Mitigating Unproductive Reasoning in Large Reasoning Models
Qingjie Zhang, Yujia Fu, Yang Wang, Liu Yan, Tao Wei, Ke Xu, Minlie Huang and Han Qiu 28532

TabSieve: Explicit In-Table Evidence Selection for Tabular Prediction
Wang Yongyao, Ziqi Miao, Lu Yang, Haonan Jia, Wenting Yan, Chen Qian and Lijun Li . . 28555

On the Robustness of Knowledge Editing for Detoxification
Ming Dong, Shiyi Tang, Ziyan Peng, Guanyi Chen and Tingting He 28575

ABDATABENCH: Evaluating AI Agents for Evidence-Grounded Antibody Record Construction
Weiye Si, Juncheng Qian, Liu Litong, Meiyan Jin, Haotian Chen, Pengrui Lu, Yunze Wu, Junjie Yu, Tianlei Ying and Pengfei Liu 28594

Follow the Latent Roadmap: Navigating Revocable Decoding for Diffusion LLMs with Anchor Tokens
Yizhen Yao, Qinglin Zhu, Runcong Zhao, Xiangxiang Dai, Yanzheng Xiang, Yulan He and Lin Gui 28615

Towards Effective Structured Context Modeling for Conversational Recommender Systems via Dual-node Monte Carlo Tree Search
Jincheng Zhang, Chen Huang, Wenqiang Lei, See-Kiong Ng and Yang Deng 28635

ADORA: Training Reasoning Models with Dynamic Advantage Estimation on Reinforcement Learning
Qingnan Ren, Shiting Huang, Zhen Fang, Zehui Chen, Lin Chen, Lijun Li and Feng Zhao . 28660

Cube Bench: A Controlled Diagnostic for Spatial and Sequential Reasoning in Multimodal Large Language Models
Dhruv Anand and Ehsan Shareghi 28678

Towards a Mechanistic Understanding of Propositional Logical Reasoning in Large Language Models
Danchun Chen, Qiyao Yan, Chenpeng Wang and Liangming Pan 28708

LadderEdit: Edit-Level Residual Compression for Memory-Efficient Lifelong Editing of LLMs
Xiaobing Yu, Peijie Qiu, Jin Yang, Xuanzhao Dong, Weiwei Ma, Zhaoqi An, Xiaoqi Zhao and Xiaofeng Liu 28738

ToolLoop: Closed-Loop Tool-Use Data Synthesis via Decomposed Generation and Dynamic Self-Feedback
Min Zeng, Yuzhou Liu, Zhenyu Cao, Hanxiu Chen, Heng Li, Caiquan Liu, Yafei Wen and Xiaoxin Chen 28765

LatentEvolve: Self-Evolving Test-Time Scaling in Latent Space
Fanci Meng, Guibin Zhang, Zherui Li, Kun Wang, Zhenfei Yin and Shuicheng Yan 28780

Can Retrieval Heads See Images? Multimodal Retrieval Heads in Long-Context Vision-Language Models
Aaron Branson Cigres Li, Zhaowei Wang, Yu Zhao, Yiming Du, Haobo Li, Xiyu Ren, Ginny Wong, Simon See, Lishu Luo, Haodong Duan, Pasquale Minervini and Yangqiu Song 28801

PN-Mix: Pair-Norm Guided Mixed-Precision Key Cache Quantization for Efficient Long-Output LLM Inference
Gil Jae Song and Sungjoo Yoo 28819

Meta-Prompt Optimization for LLM-Based Sequential Decision Making
Mingze Kong, Zhiyong Wang, Yao Shu and Zhongxiang Dai 28842

Holmes-SQL: Benchmarking LLM Agents on Evidence-Aware Investigative Reasoning over Structured Databases
Hoang-An Vo and Cam-Tu Nguyen 28868

AR-Bench: Benchmarking Legal Reasoning with Judgment Error Detection, Classification and Correction for Chinese Criminal Cases
Yifei Li, Richong Zhang, Wanyu Tu, Zhijie Nie, Haokun Luo and Chuantao Yin 28894

Understanding Reliability in LLM-based Human Behavior Simulation
Pei Wang, Lei Wang, Yuanzi Li and Xu Chen 28919

Beyond Leaderboards: A Frozen-Rubric Contract for Position-Invariant Winning-Argument Detection
Zida Yang, Weiqing Zeng, Guquan You and Jiexin Zheng 28937

Tracing the complexity profiles of different linguistic phenomena through the intrinsic dimension of LLM representations
Marco Baroni, Emily Cheng, Iria de-Dios-Flores and Francesca Franzon 28959

MedInsightBench: Evaluating Medical Analytics Agents Through Multi-Step Insight Discovery in Multimodal Medical Data
Zhenghao Zhu, Chuxue Cao, Yuanfeng Song, Xing Chen, Caleb Chen Cao, Sirui Han and Yike Guo 28984

Is Peer Review Really in Decline? Analyzing Review Quality across Venues and Time
Ilia Kuznetsov, Rohan Kumar Nayak, Alla Rozovskaya and Iryna Gurevych 29017

Not All Queries Are Equal: Query-Type-Aware Mixture-of-Experts for Temporal Knowledge Graph Forecasting
Rong Ma, Bo Ma, Yating Yang, Xi Zhou, Lei Wang, Rui Dong, Ahtamjan Ahmat and Zhen Wang 29057

Beyond Bag-of-Words: Generative Structural Regularization for Dense Table Retrieval
WeiJie Liu, Shanshan Chen and Fang Kong 29079

Distributional Stabilization for Zero-Shot Relation Extraction
Yongxin Liu, Huaping Zhang, Qiuchi Li, Chunxiao Gao, Haocheng Lv, Yan Ruohao, Baohua Zhang and Lei Li 29095

EgoVoice: Proactive Spoken Assistance from Egocentric Multimodal Streams
Heeseung Kim 29112

KItCAT: Knowledge Injection via Input Corruption for Auto-regressive Training
Meghanadh Pulivarthi, Kushagra Bhushan, Vineet Kumar, Gaurav Pandey, Jaydeep Sen, Dinesh Raghu, Sachindra Joshi and Yatin Nandwani 29137

SSPC: Semantic-Structural Propagation Calibration for Multimodal Fake News Detection
Zejiang He, Yun Luo, Zhen Huang, Zhiliang Tian, Jingyuan Huang, Shanshan Liu, Yanzhou Ma, Zhaoning Zhang and Dongsheng Li 29155

Omnilingual MT: Machine Translation for 1,600 Languages
Belen Alastruey, Niyati Bafna, Andrea Caciolai, Kevin Heffernan, Artyom Kozhevnikov, Christophe Ropers, Eduardo Sánchez, Charles-Éric Saint-James, Ioannis Tsiamas, Xiang Tony Cao, Chierh Cheng, Joe Chuang, Paul-Ambroise Duquenne, Mark Duppenthaler, Cynthia Gao, Pere-Lluís Huguet Cabot, João Maria Janeiro, Jean Maillard, Gabriel Eduardo Mejia-Gonzalez, Holger Schwenk, Edan Toledo, Arina Turkatenko, Albert Ventayol-Boada, Ravi Arikapudi, Nathanial Paul Ekberg, Rashel Moritz, Alexandre Mourachko, Surya Parimi, Mary Williamson, Shireen Yates, David Dale and Marta R. Costa-jussà 29177

CoMerge: Conflict-Driven Preference Optimization for Multi-Task Model Merging
Mingjie Zheng, Zihao Chen, Wenqing Chen, Weile Yuan, Zhixuan Chu, Jianxing Yu and Zibin Zheng 29201

Personalization over Generalization: History-Aware Sentence-Level AI Detection in Collaborative Writing
Rizky Adi, Yoshimi Suzuki and Fumiyo Fukumoto 29217

Imag-Eval: a language-grounded framework for interpretable Text-to-Image instruction following evaluation
Ibrahim Mohamed Serouis and David Jaramillo Duque 29236

Do Benchmarks Measure the Same Thing? Evaluation-Free Benchmark Redundancy Analysis with SAE
Yaoning Wang, Jiahao Ying, Yubo Ma, Yu Wan, Dayiheng Liu, Yixin Cao and Yu-Gang Jiang 29255

How Far Are AI Scientists from Changing the World? A Survey of Autonomous Scientific Discovery
Qiujie Xie, Yixuan Weng, Minjun Zhu, Junying Fang, Fuchen Shen, Shulin Huang, Zhen Lin, Zhouyang Wang, Shichen Li, Jiahui Zhou, Zilan Mao, Sifan Liu, Zhenhao Liu, Dalv Yin, Fanjunduo Wei, Zhiyuan Ning, Zijie Yang, Linyi Yang, Jian Wu and Yue Zhang 29272

MENTOR: Reinforcement Learning via Flexible Teacher-Optimized Rewards for Tool-Use Distillation
ChangSu Choi, Hoyun Song, Dongyeon Kim, Minkyung Cho, WooHyeon Jung, SunJin Park, NohHyeob Bae, Seona Yu and KyungTae Lim 29289

AP-GRPO: Anchor-Gated Phonetic Alignment with Policy Optimization for Pathological Speech Reconstruction
Pengfei Zhang, Hoang H Nguyen, Yutong Song, Wenjun Huang, Tahmid Imtiaz Imu, Henry Peng Zou, Jiang Wu, Honghui Xu and Amir M. Rahmani 29320

Towards AI-Assisted Research Writing: Benchmarking LLMs for AI/ML Introduction Generation
Krishna Garg, Firoz Shaik, Sambaran Bandyopadhyay and Cornelia Caragea 29341

DVD: A Robust Method for Detecting Variant Contamination in Large Language Model Evaluation
Renzhao Liang, Jingru Chen, Bo Jia, Bo Deng, Chenggang Xie, Yidong Wang, Ke Jin, Xin Wang, Linfeng Zhang and Cunxiang Wang 29366

Want Better Synthetic Data? Steer It: Activation Steering for Low-Resource Language Generation
Jan Cegin, Daniil Gurgurov, Yusser Al Ghussin and Simon Ostermann 29385

Canonical Color as a Lens into Concept Decodability in Vision Encoders and VLMs
Xiaofu Chen, Stella Frank and Yova Kementchedjhieva 29413

How to Attack Stealthily? A Letter Frequency-based Backdoor Attack Against LLMs Bounded by Logit Fluctuation
Ruyi Zhang, Songlei Jian, Yusong Tan, Heng Gao, Jianfeng Zhang and Haifang Zhou 29425

LoMaR: Local Margin Preservation via Residual Subspace Adaptation for Noisy Prompt Robustness in Small Language Models
Jin Hong, Jinghui Zhang, Xuwei Qian, Hongjie Lin, Jiahui Jin and Ding Ding 29444

HEAL: Anchored Hyperspherical Equilibrium for Multilingual Safety Alignment
Wenwei Chen, Hongliang Chen and Kun Zhan 29463

SaFeR-Steer: Evolving Multi-Turn MLLMs via Synthetic Bootstrapping and Feedback Dynamics
Haolong Hu, Hanyu Li, Tiancheng He, Huahui Yi, An Zhang, Qiankun Li, Kun Wang, Yang Liu and Zhigang Zeng 29473

From AI Reviewers to Evidence Assistants: Quantifying the Human-AI Responsibility Boundary in Peer Review
Zhouyang Wang, Qiujie Xie, Minjun Zhu, Shichen Li, Shulin Huang, Han Cui, Yiran Ding, Panzhong Lu, Zhenhao Liu, Fuchen Shen, Junshu Pan, Dalv Yin, Ke Sun, Zhiyuan Ning, Yixuan Weng, Peifeng Li and Yue Zhang 29494

Size Sound Symbolism Survives Genealogy Across 27 Languages
Anika Sharma, Emma Wrenn, Tianyi Niu and Shashank Srivastava 29520

REINS: Refusal-Enhanced Inhibitory Steering with Sparse Autoencoder Features
Kai-Xuan Ding, Hao-Xiang Xu, Ji-Hua Peng, Zi-Qi Chen, Jiaqi Wang and Zhen-Hua Ling 29541

Learning to Remember: End-to-End Training of Memory Agents for Long-Context Reasoning
Kehao Zhang, Shangtong Gui, Sheng Yang, Chen Wei and Yang Feng 29575

Debias-SparseGPT: Bias-Aware Pruning for Large Language Models
Irina Proskurina, Guillaume Metzler, Antoine Gourru and Julien Velcin 29600

3D-MoE: Towards Spatial Intelligence with Mixture-of-Experts for 3D Reasoning and Action Generation
Yueen Ma, Zenglin Xu and Irwin King 29624

LLM Reasoning for Machine Translation: Synthetic Data Generation over Thinking Tokens
Armel Randy Zebaze, Rachel Bawden and Benoît Sagot 29644

RATF: Robust Adaptive Tri-Encoder Framework for Fake News Detection against Style-Confusion Attacks from LLMs
Haoda Zhang, Zhenhua Tan and Ziwei Cheng 29683

GUI-PRA: Process Reward Agents for GUI Tasks
Tao Xiong, Xueyu Hu, Yurun Chen, Yuhang Liu, Changqiao Wu, Pengzhi Gao, Wei Liu, Jian Luan and Shengyu Zhang 29695

Peeling the Matryoshka: Extracting Ontological Structure from Retrieval Embeddings
Eunchan Kim and Inwhee Joe 29717

Not All Evidence Is Equal: Source-Sensitive Calibration in Large Language Models
Sercan Karakaş and Yusuf Şimşek 29730

KnowSim: Evaluating Information Calibration in LLM Assistants with User Simulators that Learn
Yoonjoo Lee, Hyoungwook Jin, Tae Soo Kim, Shaoyang Zhang, Philippe Laban and Q.Vera Liao 29750

HoliTok: A Continuous Holistic Tokenization with Robust Dual Capabilities of Speech Generation and Understanding
Bohan Li, Lianshi, Hankun Wang, Yiwei Guo, Yu Xi, Zhihan Li, Da Zheng, Colin Zhang and Kai Yu 29781

MaS-VQA: A Mask-and-Select Framework for Knowledge-Based Visual Question Answering
Xianwei Mao, Kai Ye, Sheng Zhou, Nan Zhang, Haikuan Huang, Bin Li and Jiajun Bu . . . 29796

LLM Fine-tuning with an Unconstrained Fine-tuned Teacher
Jianhao Zhang, Ou Wu and Yi Du 29812

Benchmarking Patent Drafting from Inventor-Style Disclosures
Lekang Jiang, Wenjun Sun and Stefan Goetz 29840

SWE-MEM: Learning Adaptive Memory Management for Long-Horizon Coding Agents
Shuzheng Gao, Wenhao Zeng, Zhaojian Yu, Jianqiao Wangni, Chaozheng Wang, Kai Cai, Shilin He and Michael R. Lyu 29855

Self-Corrected Speech Reasoning: Diagnosing and Improving Condition Updating via PACER
Zijie Zeng, Liwen Shi, Yunji, Bowen Wang, Nanjie Li, Xiaoyong Guo, Zeyu Chen, Xinzheng Sun, Xuyang Xiong, Haoyang Wu and Sheng Zhang 29882

I Am Multitudes: A Survey and Roadmap for Rethinking Identity in NLP
Arianna Muti, Elisa Bassignana and Debora Nozza 29912

Bridging Text and Structure via Latent Query Refinement for Knowledge Graph Completion
Chunmiao Yu, Zikang Wang, Tianyi Luo, Jiaojiao Wang and Zhidong Cao 29932

OptiSkill: A Hierarchical and Evolving SkillBank for LLM-Based Optimization Modeling
Ruiqing Zhao, Rui Liu, Yuan Zuo, Huarong Zhang, Xiao Han and Junjie Wu 29946

Reviewing the Reviewer : LLM-Assisted Reviewer Feedback Generation for Guideline Compliance
Sukannya Purkayastha, Qile Wan, Anne Lauscher, Lizhen Qu and Iryna Gurevych 29968

Can Humans Dream of Electric Sheep? Human-Written Samples for Fine-Grained Vision-and-Language Hallucination Benchmarking
Timothee Mickus, Claudio Savelli, Eduardo Calò, Emilio Raimond, Stella Frank, Hengyu Luo, Flavio Giobergia, Vincent Segonne, Chuyuan Li, Aman Sinha, Lorenzo Vaiani, Jörg Tiedemann and Raúl Vázquez 30009

TriggerBench: Investigating Prospective Memory for Large Language Models
Tianhua Zhang, Xinjiang Wang, Qianxi Zhang, Qi Chen, Kun Li, Yaoqi Chen, Dingdong Wang, Helen M. Meng and Yan Lu 30034

TORe: Rehearsal-Free Continual Learning for Table QA via Operation Recipe Reconstruction
Jinyu Zhang, Ruiheng Liu and Yu Zhang 30069

Revisiting Reinforcement Learning with Verifiable Rewards from a Contrastive Perspective
Feng Zhang, Xinhong Ma, Ziqiang Dong, XI Leng, Jianfei Zhao, Xin Sun, Yang Yang and Guanjun Jiang 30096

RePro: Proof-Verified Benchmark Rewriting for Reliable Evaluation of LLM Mathematical Problem Solving
Xiyuan Zhou, Zhuoqi Li, Xinlei Wang, Yirui He, Yuhao Wu, Yuheng Cheng, Yan Xu, Junhua Zhao and Jinjin Gu 30112

How Much is a Human Right Worth? ECtHR-NPD: A Benchmark for Predicting Non-Pecuniary Damage Awards
Yanyi Pu, Damian Gonzalez-Salzberg, Zheng Yuan and Nikolaos Aletras 30138

Diagnosing LLM Fragility: An Empirical Study on Reasoning Entropy and Malicious Outputs
Fabio Mercorio, Daniele Potertì and Andrea Seveso 30166

Phrase-Localized Language-Contrastive Guidance: Training-Free Localized Accent Control for Code-Switching Text-to-Speech
Che Hyun Lee, Sangkwon Park, Donghun Kang, Dongwook Lee, Youngho Cho, Heeseung Kim and Sungroh Yoon 30185

Refuse without Refusal: A Structural Analysis of Safety-Tuning Responses for Reducing False Refusals in Language Models
Minji Kim and Hyounghun Kim 30203

Forgetting Across Modalities: Unlearning in Multimodal Large Language Models
Yan Li, Xingrui Yu, Zhiqiang Wang, Zhenyi Wang and Ivor Tsang 30241

Knowing What Not to Answer: Selective Non-Compliance in Vision-Language Models
Minji Kim, Jihyoung Jang and Hyounghun Kim 30262

CAME: Company-Aware Evidence-Memory Experts for Interpretable Quarter-Ahead Revenue Forecasting
Ya-Wen Wu, Meng-Fen Chiang, Kuang-Da Wang and Wen-Chih Peng 30305

Where Does an LLM Decide What's Right? Sparse Normative Circuits Across Languages and Model Families
Zhitao Huang, Qiang Huang, Wei Zhai, Lemao Liu, Yingmao Tang and Dejing Dou 30329

Auditing Chinese Web-scale Corpora via Sampled BPE Token Statistics
Qingjie Zhang, Ziqi Tang, Jie Zhang, Gelei Deng, Jinfeng Li, YueFeng Chen, Yitong Yang, Hui Xue, Tianwei Zhang and Han Qiu 30352

Cross-Version Δ-Transfer of QLoRA-Trained Adapters
Yoshiki Kubota, Hirotada Honda, Takashi Sano and Shugo Nakamura 30380

DEPPO: Dual Experience Pool Policy Optimization for Long-Horizon Reinforcement Learning
Haosen Dong, Jiaxin Guo, Muyang Yan, Ye Yuan, Mengxi Jia and Hao Sun 30402

When Fairness Evolves: Value Orientation and Evolution in Multi-Agent Resource Allocation
Song Gao, Zengqing Wu, Yu-Xuan Qiu, Chuan Xiao, Ruiqi Xu and Jianbin Qin 30419

Ask, Answer, or Act: Learning Interaction Modes for Tool-Use Agents
Tianyu Fu, Yangrui Xie, Yunpu Ma and Wanyuan Wang 30447

Efficient Listwise Reranking with Compressed Document Representations
Hervé Déjean and Stéphane Clinchant 30474

Colla-Q: Toward Collaborative Experts in MoE Quantization via Minimax Precision Balancing
EunJu Shin and Jongbin Ryu 30486

In-Place Feedback: Reliable Refinement for Multi-Turn Expert-LLM Collaboration
Youngbin Choi, Minjong Lee, Saemi Moon, Seunghyuk Cho, Chaehyeon Chung, MoonJeong Park and Dongwoo Kim 30503

AdaPLD: Adaptive Retrieval and Reuse for Efficient Model-Free Speculative Decoding
Runheng Liu, Jincheng Xie, Wen Hu, Xingchen Xiao and Heyan Huang 30546

SpecTr-GBV: Multi-Draft Block Verification Accelerating Speculative Decoding
Yijun Lin, Jin-Hao Sheng, Qingyue Cai and Feng Zhou 30562

I^3CL: Restoring Interactive Capabilities to Implicit In-Context Learning
Hao Wu, Bin Shi, Kai Sun, Aoxiang Zhou, Ziqi Liu, Zhiqiang Zhang, Jun Zhou, Wenhan Zeng, Peng Liu and Bo Dong 30579

GutenBench: A Taxonomy-Based Dataset for Evaluation of Information Retrieval Models
Marcio Monteiro, Denys Senkin, Abhilasha Ravichander, Marius Kloft and Sophie Fellenz30596

KairosAgent: Agentic Time Series Forecasting with Fused Semantic Reasoning
Kun Feng, Ziwei Shan, Yuchen Fang, Yiyang Tan, Sihan Lu, Shuqi Gu, Xingyu Lu, Lintao Ma and Kan Ren 30614

INCLUDE-2.0: Disentangling Knowledge Gaps in Multilingual LLM Evaluation
Angelika Romanou, María Grandury, Clara Meister, Negar Foroutan, Jabez Magomere, Anna Sotnikova, Hitesh Laxmichand Patel, Akshay Kumar Dubey, Alfonso Amayuelas, Amit Agarwal, Arion Das, Carlos Arriaga, Cheng Yu, Dalia Ali, David Quispe-Castillo, Fenal Ashokbhai Ilasariya, Francisco-Javier Rodrigo-Ginés, Gonzalo Martínez, Gouthami Vadithya, Guido Ivetta, Javier Conde, Jorge Vallego, Juan J. Calderón G., Karan Gupta, Marcos J Gomez, Marvin M. Agüero-Torales, Mithil Bangera, Naira Paola Arnez Jordan, Oscar M. Cumbicus-Pineda, Priyaranjan Pattnayak, Reewos Erwin Talla Chumpitaz, Sahil Aggarwal, Sahil Raj, Shruti Jalan, Srikant Panda, Susana Zhou, Yeshil Bangera, Orestis Papakyriakopoulos, Scott A. Hale, Shamsuddeen Hassan Muhammad and Antoine Bosselut 30643

Targeted Token-Level Unlearning for Repairing Text-to-SQL Models
Jinyu Zhang, Ruiheng Liu and Yu Zhang 30667

JarvisGUI: Towards Cross-Device GUI Agents with Dynamic Task Composition
Zixiang Chen, Yuheng Lu, Zihao Cheng, Zeming Liu, Jizeng Bai, Ziye Huang, Zhiyin Lin, Zihan Li, Yuhang Guo, Yunhong Wang and Haifeng Wang 30688

MARS: What Retrieval Signals Are Hidden in Multimodal Large Language Models for Text-Video Retrieval?
Uicheol Jung, Juyoung Hong, Geuntaek Lim and Yukyung Choi 30716

BEAR: Budgeted Evidence Allocation for Multi-Document Reasoning
Lin Sun, Linglin Zhang, Jingang Huang, Change Jia, Zhengwei Cheng and Xiangzheng Zhang 30731

ToToFilter: Topic- and Topology-Aware Noise Filtering Framework over Medical KGs for Evidence-Based Medical Q&A
Yufeng Wang, Weijie Chen, Chao Chen, Xingfeng Liu, Yizhe Yang, Zhihao Chen, Qiming He, Yubo Chen, Kang Liu, Xuefeng Wang, Xiangwen Liao and Jingfeng Liu 30750

Breaking Temporal Ambiguity: Linguistic Temporal Knowledge Transfer for Weakly-Supervised Action Localization
Zhilin Li, Shining Wu, Yixin Zhang, Chuandong Cheng and Zilei Wang 30771

ELITE: Evolving Localized and Informative Temporal Entity alignment
Jiayun Li, Wen Hua, Shiqi Fan, Yuchuan Zhao, Wei Jiang, Haiyang Jiang and Xue Li 30788

Linear Probing Provides Robust and Efficient Detection of Machine-Generated Text
Gerrit Quaremba, Hanqi Yan, Elizabeth Black, Denny Vrandečić and Elena Simperl 30805

EHR-Complex: Benchmarking Medical Agents for Complex Clinical Reasoning
Yitong Qiao, Lei Liu, Yue Shen, Jian Wang, Jinjie Gu, Zhixuan Chu and Kui Ren 30828

I Know What You Meme, Even If it Emerged Today: Understanding Evolving Memes through Open-World Knowledge Acquisition
Shanhong Liu, Rui Cao, Pai Chet Ng and De Wen Soh 30853

AnyCap: Omni-Modal Captioning with Instruction Alignment
Yiming Ren, Zhiqiang Lin, Yu Li, Meng Gao, Junjie Wang, Zicheng Lin, Jifeng Dai, Yujiu Yang, Wenhai Wang and Ruihang Chu 30869

QReason: Query-Focused Decoupled Chain-of-Thought for Efficient Passage Reranking
Yang Zhang, Wenhan Liu, Qiannan Zhu, Mingming Li and Yuanfei Huang 30902

C³T: Counterfactual Causal Reasoning for Sentiment Shifts in Social-Media Conversation Trees
Rifat Rafiuddin and Atriya Sen 30926

FullFront: Benchmarking MLLMs Across the Full Front-End Engineering Workflow
Haoyu Sun, Huichen Will Wang, Jiawei Gu, Linjie Li and Yu Cheng 30948

Multi-Specialist LLM Relay System for Competitive Programming
Andrei Mikhailov, Mikhail Burtsev and Alsu Sagirova 30995

EmoStance: Response-Side Affective-Orientation Control for Empathetic Response Generation via Emoji Weak Supervision
Ziyuan Jin, Yuxuan Ge and Zheng Tian 31008

ESGPT: A Data-Centric ESG Language Model with Unified Instruction and Benchmark Resources
Jaeyoung Lee and Misuk Kim 31043

MedVision: Benchmarking Quantitative Medical Image Analysis
Yongcheng Yao, Yongshuo Zong, Raman Dutt, Yongxin Yang, Sotirios A. Tsaftaris and Timothy Hospedales 31065

A²Agent: Action-Aware Reinforcement Learning for Repository-Level Code Localization Agents
Doyeon Kim, Suyoung Bae, Yumin Lee and Jee-Hyong Lee 31103

MemPro: Agentic Memory Systems as Evolvable Programs
Qingshan Liu, Guoqing Wang, Wen Wu, Jingqi Huang, Xinqi Tao, Dejia Song, Jie Zhou and Liang He 31124

How Reliable are AMR Parser Outputs? Uncertainty Estimation for AMR Parsing
Anuj Kumar, Shintaro Ozaki, Yamuna Prasad, Katsuhiko Hayashi, Hidetaka Kamigaito and Taro Watanabe 31144

It's Not RoPE that Creates Sinks: The Role of Self-Concentration and Value-Non-Mixing in Attention
Raito Kiya, Satoki Ohashi, Kosuke Sato, Go Kamoda, Ryosuke Takahashi, Yuji Yamamoto, Daiki Shiono, Keisuke Sakaguchi and Goro Kobayashi 31160

LoopCD: Loop-wise Contrastive Decoding for Improving Reasoning in Looped Language Models
Byeongho Yu, Junhyuk So and Eunhyeok Park 31185

MedSPA: Self-planning Medical Agents
Luu Van Huy, Bowen Wang, Fumio Okura and Yuta Nakashima 31200

When the Strongest Teacher Is Not the Best Teacher: Student-Centric Answer Selection
Zhengyu Hu, Zheyuan Xiao, Linxin Song, Fengqing Jiang, Yuetai Li, Zhihan Xiong, Junhao Lin, Yue Liu, Yao Su, Lijie Hu, Kaize Ding, Teng Xiao and Radha Poovendran 31221

Don't Guess, Ask Wisely: Benchmarking and Training Proactive Elicitation in Open-Domain LLM Conversations
Rongyu Zhang, Mutian Bao, Jinjian Zhang, Yuetian Du, He Xu, Wei Zhou, Xiaojun Zhu, Ming Kong, Jie Liu, Linjian Mo and Qiang Zhu 31257

Beyond FLOPs: Benchmarking Real Inference Acceleration of LLM Pruning under a GEMM-Centric Taxonomy
Haozhe Hu, Hao Wu, Anhao Zhao, Longwei Ding, Peiran Yin, Yunpu Ma and Xiaoyu Shen31283

DecSelfMask: Leveraging Unlabeled Text via Self-Relevance-Guided Masking for Decoder-Only Classification
Pietro Ferrazzi, Matteo Merler, Giovanni Bonetta, Alberto Lavelli and Bernardo Magnini . 31311

VideoSearcher: Empowering Video Deep Research with Multi-Tool Agentic Reasoning via Reinforcement Learning
Zhenkun Gao, Yicheng Bao, Jinlong Peng, Xueheng Li, Suyuan Huang, Bangwei Liu, Kunquan Li, Zhenye Gan, Tao Hu, Chengjun Xie, Xuanhua He, Zhizhong Zhang, Xin Tan, Chengjie Wang and Yuan Xie 31328

AT-TTA: Dual-View Anchor-Tethered Test-Time Adaptation for Large Language Models
Yaxin Wang, Junfeng He, Zhihui Lu, Weishen Lu, Shijing Hu, Peng Chen, Ye Wang, Chao Huang, Yao Cheng and Yudong Cheng 31353

From Unverifiable Domain Text to Cross-Lingual RLVR via Code-Switched Task Synthesis
Cheonbok Park, Jihwan Kim, Shinyoung Joo, Joosung Lee, Hwiyeol Jo, Jeonghoon Kim and Jaegul Choo 31382

MARBO: Relational Belief Grounding for LLM Agents in Social Deduction Games
Yechan Hwang, Sangjun Bae, Jeongmo Kim, Sangwoo Bang and Seungyul Han 31414

Speak for Me: Giving LLMs the Situational Awareness to Participate in a Meeting
Muneeb Khan, Frederic Kirstein, Terry Ruas and Bela Gipp 31446

How Well Can LLMs Simulate Real Learner Evaluations of Educational Feedback?
Momoka Furuhashi, Kouta Nakayama, Takashi Kodama, Saku Sugawara and Kyosuke Takami 31473

PROBE: Prospective Retrieval with Oracle-Based End-Goal Distillation for Multi-Hop Question Answering
Xu Lu, Zhengqing Zang, Binhui Yao, Luwei Yang, Ruimin Chen, Zeng Wang and Haobo Wang 31494

Anchoring Speech with Semantics: A Multimodal Adapter Mechanism for Automatic Speech Recognition in Low-Resource Languages
Kuan-Tang Huang, Cheng-Yeh Yang, Chien-Chun Wang, Hung-Shin Lee, Hsin-Min Wang and Berlin Chen 31516

SMADE-IE: Sparse Multi-Agent Framework with Evidence-Driven Debate for Zero-Shot Information Extraction
Kenfeng Huang, Yi Cai, Xin Wu, Zikun Deng and LI Yuan 31532

Right for the Right Reasons? Evaluating the Structural Fidelity of LLM Emotion Explanations
Hajime Hotta and Minh-Tien Nguyen 31555

Answer Probing-Guided Search for Diverse Solution Exploration of LLMs
Yi Fang, Que Shen, Chengpeng Li, Boyi Deng, Wei Shi, Wenjie Wang, Fuli Feng, Fengli Xu and Dayiheng Liu 31571

Trip+: Benchmarking Agents in Personalized Interactive Travel Planning
Junle Chen, Wei Chen, Yehong Xu, Zhengjun Huang, Yuqian Wu, Zhoujin Tian, Kai Wang, Rui Chen and Xiaofang Zhou 31588

RIGOR: Atomic Semantic Normalization and Declarative Constraint Reasoning for Geometry Problem Solving
Yuhang Dong, Weijie Yu, Ziheng Wang, Kunhang Lv, Rui Han, Fuqi Jia, Minghao Yin and Feifei Ma 31626

Cognition on Graph: Navigating Massive Knowledge Space via Cognitive Cycles and Bidirectional Graph-Text Synergy
Gengxian Zhou, Jian Xu, Zichen Tang, Shiming Xiang, Haihong E and Cheng-Lin Liu . . . 31643

Rethinking Gradient Flow Through Frozen Blocks for Memory-Efficient Block-Coordinate Training
Yeachan Kim 31675

Beyond Sparse Rewards: A New Benchmark and Structure-Aware Graph Alignment for Micro-Drama Understanding
Yixin Qin, Shi-Zhe Chen, Zhiqi Yu, Siyuan Cheng, Tao Cheng, Jinwen Luo and Zheng Wei31686

CounterVid: Counterfactual Video Generation for Mitigating Action and Temporal Hallucinations in Video-Language Models
Tobia Poppi, Burak Uzkent, Amanmeet Garg, Lucas Resque Porto, Garin N. Kessler, YZ Yang, Marcella Cornia, Lorenzo Baraldi, Rita Cucchiara and Florian Schiffers 31712

Self-Sanitize: Cognitive-Inspired In-Place Mitigation of LLM Harmful Responses
Wenjie Fu, Huandong Wang, Chen Gao, Guoan Wan, Junyao Gao, Yong Li and Tao Jiang . 31732

Disentangling Semantics and Popularity: Adversarial Debiasing and Asymmetric Alignment for LLM-Enhanced Long-Tail Sequential Recommendation
Wenzhe Dong, Minglai Shao and Qiyao Peng 31758

InComeS: Integrating Compression and Selection Mechanisms into LLMs for Efficient Model Editing
Shuaiyi Li, Zhisong Zhang, Yang Deng, Chenlong Deng, Tianqing Fang, Hongming Zhang, Haitao Mi, Dong Yu and Wai Lam 31775

Learning User Simulators with Turing Rewards
Yingshan Susan Wang, Cedegao E. Zhang, Linlu Qiu, Zexue He, Pengyuan Li, Alex Pentland, Roger P. Levy and Yoon Kim 31797

LEVEE: Layer-wise Evaluation of Vector Encoding Effects: Bias Localization and Functional Separation across 11 LLMs
Dongmin Kim and Youhyun Shin 31845

Rethinking Pivot Programming Languages in Code Language Models
Andor Diera, Lukas Galke Poech and Matthias Tichy 31875

Extending AI for Research to the Humanities: A Multi-Agent Framework for Evidence-Grounded Scholarship
Yating Pan, Jiajun Zhang, Jun Wang and Qi Su 31893

Bridging Time Series and Language via Query-Adaptive Multi-Granularity Routing for Temporal Reasoning
Hyuna Jeon, Jaeyoung Lee and Misuk Kim 31917

Messier: A High-Resolution Corpus for Cross-Benchmark Agent Evaluation
Stefan Krsteski, Charlotte Meyer, Guillaume Allegre, Tony O'Halloran and Alexandre Sallinen 31936

Addressing Corpus Knowledge Poisoning Attacks on RAG Using Sparse Attention
Sagie Dekel, Moshe Tennenholtz and Oren Kurland 31952

HybridThinker: Efficient Chain-of-Thought Reasoning via Compressed Memory and Transient Thought Steps
Xin Liu, Runsong Zhao, Xinyu Liu, Junhao Ruan, Pengcheng Huang, Shichao Dong, Chunyang Xiao, Chenglong Wang, Changliang Li, JingBo Zhu and Tong Xiao 31974

A Target-Centric Survey of Quantization-Aware Training
Jiamin Song, Mengjie Zhao, Zijing Wang, Yongkang Liu, Qian Li, Shi Feng, Feiliang Ren, Daling Wang and Hinrich Schuetze 32001

COPILOT: Controlling Fine-grained Provider Preference in LLM Recommendations
Xiaoyu Zhang, Yimin Ding, Qingshuang Bao, Chenhao Lin, Chao Shen, Bo Yang and Yang Liu 32027

DLawBench: Evaluating LLMs Through Multi-Turn Legal Consultation
Li Zhang, Yuzhen Shi, Yiran HU, Zhang Jingwen, Wenbo Lv, Yubo Ma, Wei Wang, Rongyao Shi, Yuanyang Qiu, Xu Xinran, Yuemeng Qi, Linlin Miao, Jaromir Savelka, Yun Liu, Kevin D. Ashley, Bing Zhao, HU Wei and Lin Qu 32053

How and What to Imagine? Visual Thinking in Unified Multimodal Models for Cross-View Spatial Reasoning
Qian Yang, Ankur Sikarwar, Huy Le, Le Zhang, Zhuan Shi, Perouz Taslakian and Aishwarya Agrawal 32092

Decoupled Physical Modeling and Execution for Physics Reasoning
Ye Zhang, Xuehang Guo, Rui Pan, Pengfei Yu, Denghui Zhang, Manling Li and Qingyun Wang 32112

ParsHate: A Benchmark Dataset for Hate and Target Detection in Persian
Zahra Bokaei, Walid Magdy and Bonnie Webber 32132

LAK-LoRA: Layer-Wise Adaptive Expert Allocation and Principal Component–Based Feature Separation for Medical Multi-Task Learning
Ye Lyu, Chuan Li, Chengyu Wang, Mingyuan Fan and Cen Chen 32150

MCite-RL: Towards Reliable Multimodal RAG via Citation-enhanced Agentic Reinforcement Learning
Suifeng Zhao, Zida Liu, Xinyu Lei, Lei Sun, Jun Gao and Sujian Li 32172

EpiCurveBench: Evaluating VLMs on Epidemic Curve Digitization
Thomas Berkane and Maimuna S. Majumder 32188

Cultural Evolution Drives the Emergence of Conventions in Neural Emergent Communication
Maxime Toquebiau, Jérôme Botoko Ekila, Paul Van Eecke and Katrien Beuls 32202

F-GRPO: Don't Let Your Policy Learn the Obvious and Forget the Rare
Daniil Plyusov, Alexey Gorbatovski, Boris Shaposhnikov, Viacheslav Sinii, Alexey Malakhov, Daria Korotyshova and Daniil Gavrilov 32224

SHAPE: Structural Hypergraph Augmentation with Personalized Event Memory for Temporal Knowledge Graph Reasoning
Qing Li, Jialin Lv, Jingjing Song, Duan Fei and Zhongwen Guo 32251

Mitigating LLM Over-Refusal via Dynamic Semantic Routing Calibration
Zixuan Wang, Bingjie Zhang, He Zhao and Dandan Guo 32266

JADES: A Universal Framework for Jailbreak Assessment via Decompositional Scoring
Junjie Chu, Mingjie Li, Ziqing Yang, Ye Leng, Chenhao Lin, Chao Shen, Michael Backes, Yun Shen and Yang Zhang 32289

SARATHI: Decoupled Data-Free Probing and Adaptive Reconstruction for Structured LLM Pruning
Sankar Behera, Dhruv Singh, Satyadev Ahlawat and Yamuna Prasad 32314

Incorporating Explicit and Implicit Chain-of-Thought as Hybrid Reasoning Paradigm with Guided Latent Space Modeling
Guangyu Li, Zhekai Ge, Zhuo Chen, Zhen Zhang, Xinyu Wang, Yong Jiang and Kewei Tu 32329

MIND: An Evidence-State Decision Interface for Criteria-Grounded Psychiatric Inquiry and Diagnosis
Guoyi Li, Shihao Xu, Jiatong Ma, Zhongjiang Yao, Yunyun Han, Jianhua Chen and Yafeng Deng 32348

Object Presence Is Not Relation Evidence: Relation-Grounded Faithful Reasoning for Multimodal Large Language Models
Xiyao Dong, Yilong Chen, Xiaojin Zhang and Kun He 32379

UEmbed: Unified Sparse and Dense Multimodal Embeddings
Tingyu Song, Mingxin Li, Yanzhao Zhang, Dingkun Long, Pengjun Xie, Zhijie Nie, Yilun Zhao and Shu Wu 32396

R^3L: Reflect-then-Retry Reinforcement Learning with Language-Guided Exploration, Pivotal Credit, and Positive Amplification
Weijie Shi, Yanxi Chen, Zexi Li, Xuchen Pan, Yuchang Sun, Jiajie Xu, Xiaofang Zhou and Yaliang Li 32415

Assessing Adversarial Robustness of Latent Reasoning Models
Shaolong Chen, Ang Li, Mingjie Li and Yisen Wang 32448

Expert-Aware Causal Tracing of Factual Recall in Sparse MoE Language Models
Yuetian Lu, Ali Modarressi, Yihong Liu and Hinrich Schuetze 32466

LongNovel: A Multi-Scale Benchmark for Hallucination Detection in Long-Context Novel Summarization
Ruizhi Zhang, Jinwei Chen, Lu Xiangju, He Yan, Mo Yu, Junmin Zhu and Wei Zhang 32478

On-Policy Distillation with Curriculum Turn-level Guidance for Multi-turn Agents
Gengsheng Li, Mao Zheng, Mingyang Song, Ruiqi Liu, Tianyu Yang, Jie Sun, Qiyong Zhong, Dan Zhang, Haiyun Guo, Junfeng Fang and Jinqiao Wang 32539

Label Over Logic? How Source Cues Bias Human Fallacy Judgments More Than LLMs
Mahjabin Nahar, Nafis Irtiza Tripto, Aiping Xiong, Ting-Hao Kenneth Huang and Dongwon Lee 32556

CLASP: A Closed-Loop Behavior Alignment Framework for Social Media Agent Simulation
Laisong Li, Shanfeng Wang, Wei Xie, Zhenhua Wang, Zhenhao Luo and Baosheng Wang . 32590

MIRAGE: Reliability Assessment via Grounded Evidence for Table Reasoning
Jinlong Tian, Ruochun Jin, Shixuan Liu, Yanfang Zhou, Mengmeng Li, Haotian Wang, Xinhai Xu and Yuhua Tang 32610

CoFGraph: Interleaving Reciprocal Calibration with Confidence Feedback for LLM-Augmented Graph Learning
Feng Zhao, Tiewen Chen, Xianggan Liu, Ruilin Zhao and Yifan Song 32624

SAGE: How Well Do LLMs Align with Classical Chinese Philosophy?
Yuying Zhao, Sirui Chen, Yixin Liu, Xuemei Jia, Lingxi Min and Zheng Wang 32636

GRASP: Geometry-aware Residual Alignment for Scalable Pretraining Data Attribution
Yue Min, Ruining Chen and Yujun Li 32663

AtomMem: Learnable Dynamic Agentic Memory with Atomic Memory Operation
Yupeng Huo, Yaxi Lu, Zhong Zhang, Haotian Chen and Yankai Lin 32684

AURA: Intent-Directed Probing for Implicit-Need Surfacing in Situated LLM Agents
Yang Li, Jiaxiang Liu, Jiang Cai and Mingkun Xu 32700

AVSCap: Orchestrating Audio-Visual Synergy for Omni-modal Video Captioning
Yanghai Wang, Jiahao Wang, Jiafu Tang, Yuanxing Zhang, Zhe Cao, Hanyan Bian, Zijie Zhang, Weiliang Luo, Zhiyu Pan, Zixuan Dong, Jiaheng Liu and Zhaoxiang Zhang 32731

MOAQ: A Multi-Scale Outlier-Aware Framework for Extreme Quantization of Large Language Models
Jiaxi Tan, Xudong Li, Yan Zhang, Jiayi Ji, Xiawu Zheng, Yunhang Shen, Ke Li and Liujuan Cao 32762

Minimal Grammars as Supervision for DSL Generation
David Vella Zarb, Geoff Parks and Timoleon Kipouros 32782

TAC-CoT: Tool-Empowered Implicit Chain-of-Thought via Interleaving Explicit Tokens
Juseon Do, Sangmin Bae, Jeesu Jung and Hwanjun Song 32803

Shaped by Values: Language Through the Lens of Culture and Morality.
Neele Falk, David Jurgens and Gabriella Lapesa 32823

Debating the Unspoken: Role-Anchored Multi-Agent Reasoning for Half-Truth Detection
Yixuan Tang, Yirui Zhang, Hang Feng and Anthony Kum Hoe Tung 32858

Perceive Before Reasoning: A Pre-Reasoning Perception Framework for Efficient and Reliable Proactive Mobile Agents
Zhijie Ding, Weinan Hong, ZiCheng Zhu, Lei Li, Dezhi Kong, Wang Hao, Peng Zhou, Xuchu Jiang and Jiaming Xu 32873

Ask, Condition or Abstain: Reinforcement Learning for Missing-Premise Reasoning
Yongqi Tong, Zhenyu Zhang, Zimi Liu, Kewei Fu, Mingli Song, Haofei Zhang, Junshao Zhang, Hong Zhu, Jiang-Ming Yang, Xin Zhang and Jianshe LI 32895

NLP Meets the World: Toward Improving Conversations With News Media About Natural Language Processing Research
Shomir Wilson 32910

ROT: Rotating Hidden States towards Contextual Vectors for Hallucination Mitigation in LVLMs
Yijing Du, Xiangcheng Zhan and Shuo Yang 32918

MA-Mesh: Resilient Multi-Agent Collaboration via Role-based Self-Healing
Junyong Lee, Woo Kyung Kim and Honguk Woo 32933

Resource Consumption Threats in Large Language Models: A Survey
Yuanhe Zhang, Xinyue Wang, Zhican Chen, Weiliu Wang, Zilu Zhang, Zhengshuo Gong, Zhenhong Zhou, Kun Wang, Li Sun, Yang Liu and Sen Su 32962

PRICE: Preemptive Confidence Estimation in LLMs via Internal Representations
Jonghak Jang and Misuk Kim 32993

DIANOIA: Diagnostic Decomposition and Joint Optimization for Multi-Agent Reasoning
Yiming Yang, Zhuoyuan Li, Fanxiang Zeng, Hao Fu and Yue Liu 33019

Does Faithfulness-Guided Alignment Hurt Accuracy? Unlocking Accurate and Faithful Post-Retrieval Reasoning
Yu Liu, Wenxiao Zhang, Diandian Guo, Cong Cao, Fangfang Yuan, Qiang Sun, Yanbing Liu, Jin B. Hong and Zhiyuan Ma 33054

Can Large Language Models Handle Discourse Particles? A Case Study of Colloquial Malay
Mariah Al Giptiah Binte Yusoff, Jakin Tan, Bocheng Chen, Guangliang Liu and Xi Chen . 33073

Training Matryoshka Mixture-of-Experts for Elastic Inference-Time Expert Utilization
Yaoxiang Wang, Qingguo Hu, Yucheng Ding, Ruizhe Wang, Yeyun Gong, Jian Jiao, Yelong Shen, Peng Cheng and Jinsong Su 33093

CirCA: Aligning Credit Assignment with Circuit Semantics via Atomic Logic Blocks for RTL Generation
Lei Huang, Rui Zhang, Jiaming Guo, Yang Zhang, Di Huang, Zidong Du, Xing Hu, Qi Guo and Yunji Chen 33112

SumRank: Aligning Summarization Models for Long-Document Listwise Reranking
Jincheng Feng, Wenhan Liu and Zhicheng Dou 33129

ScalePRM: Training Process Reward Models by Scaling Verification Compute Without Ground Truth
Salman Rahman, Sruthi Gorantla, Arpit Gupta, Swastik Roy, Nanyun Peng and Yang Liu . 33157

Identifying and Transferring Reasoning-Critical Neurons: Improving LLM Inference Reliability via Activation Steering
Fangan Dong, Zuming Yan, Xuri Ge, Zhiwei Xu, Mengqi Zhang, Xuanang Chen, Ben He, Xin Xin, Zhumin Chen and Ying Zhou 33179

AnaBench for Enhancing Scientific Table & Figure Analysis
Xuehang Guo, Zhiyong Lu, Tom Hope and Qingyun Wang 33199

Fractional Rotation, Full Potential? Investigating Performance and Convergence of Partial RoPE
Mohammad Aflah Khan, Krishna P. Gummadi, Manish Gupta and Abhilasha Ravichander 33270

Ontology-Based Knowledge Graphs for Industrial Standards with Scope-Aware Logical Rule Structuring
Jiin Park, Hyuna Jeon, Yoonseo Lee and Misuk Kim 33282

LoCoDec: Robust Decoding for LLM-based Sequence-Structured Prediction
Chang Liu, Peiyan Wang, Yubo Feng, Libang Wang, Feiyan Jiang and Yan Song 33312

RACE: Scalable Statistical Estimation of Functional Consistency in LLM Neurons
Runyu Wang, Bo Liu, Xiaxin Zhang, Yu Han, Jiawei Cao, Xiaoye Zhang, Zhe Zhang, Yifan Yang and Peng Ping 33331

Consensus Group Relative Policy Optimization for Text Generation
Yuki Ichihara, Yuu Jinnai, Kaito Ariu and Eiji Uchibe 33354

Data Efficient Sample Selection for In-Context Learning
Venktesh V, Cem Levi and Avishek Anand 33387

MultiHashFormer: Hash-based Generative Language Models
Huiyin Xue, Atsuki Yamaguchi and Nikolaos Aletras 33409

Bridging the Temporal Gap: A New Benchmark and Dynasty-aware Adaptation Framework for Ancient Chinese WSD
Yuzhi Liang, Xinrong Zhu, Lixiang Ma and Zhixiong Wan 33427

Beyond Retraining-Free MoE Compression: A Cost-Normalized Study of Post-Compression Adjustment
Sieun Hyeon and Jaeyoung Do 33444

Beyond Conditional Computation: Retrieval-Augmented Genomic Foundation Models with Gengram
Huinan Xu, Xuyang Feng, Junhong Chen, Junchen Liu, Kaiwen Deng, Kai Ding, Shengning Long, Jiaxue Shuai, Zhaorong li, Shiping Liu, Guirong Xue and Zhan Xiao 33480

Edit Knowledge, Not Just Facts via Multi-Step Reasoning over Background Stories
Ya Gao, Kalle Kujanpää, Pekka Marttinen, Harri Valpola and Alexander Ilin 33512

IterCAD: An Iterative Multimodal Agent for Visually-Grounded CAD Generation and Editing
Tao Hu, Jiaxin Ai, Licheng Wen, Xueheng Li, Shu Zou, Siqi Li, Nianchen Deng, Xinyu Cai, Hongbin Zhou, Pinlong Cai, Daocheng Fu, Yu Yang, Hairong Zhang, Botian Shi and Xuemeng Yang 33560

Arabic Morphosyntactic Tagging and Dependency Parsing with Large Language Models
Mohamed Adel, Bashar Alhafni and Nizar Habash 33582

Beyond Token Positions: Safety Alignment Across Denoising Steps in Diffusion Language Models
Guoli Wang, Haonan Shi, Tu Ouyang and An Wang 33605

Modular Representation Steering for Multi-Attribute LLM Alignment
Chaojian Shi, Renhong Huang, Xulan Qiu, Jiawei Zhou and Yunsong Zhou 33619

When Graphs Whisper: Graph Structure Serialization for Embedding Alignment in Cold-Start Recommendation
Yujing Zhang, Huachi Zhou, Ruochen Liu, Qinggang Zhang, Luyao Zhuang, Hao Chen, Zijin Hong, Chang Yang and Xiao Huang 33642

Lost Between Scripts: Recovering Arabic Words from Noisy Learner Orthographies
Mohamed Adel, Nisa Shahid and Nizar Habash 33661

Counter-GEO-Bench: Evaluating Defenses Against Information-Distorting Generative Engine Optimization
Bing Zheng, Zongyao Zhao and Wenming Yang 33673

Repetition over Diversity: High-Signal Data Filtering for Sample-Efficient German Language Modeling
Ansar Aynetdinov, Patrick Haller and Alan Akbik 33690

RepoNav: From Snippet Retrieval to File-Centered Repository Navigation for Code Agents
Hongzheng Chai, Jiakun Li, Hongyue Yu and Yuan Yuan 33711

LoRA-Patch: Defending Against Backdoored LoRA Adapters via Directional Steering
Shengyan Huo, Peihai Jiang, Xixiang Lyu and Yige Li 33727

SymboLLM-FE: LLM-Accelerated Symbolic Regression for Automated Feature Engineering on Tabular Data
Zi-Jian Cheng, Ziyi Jia, Zhi Zhou, Yu-Feng Li and Lan-Zhe Guo 33744

When LLMs Say No: Diagnosing and Mitigating Conservative Bias in Authorship Verification
Nicholas Wlezien and Rakesh M Verma 33771

A Manifold-Aware Topic Modeling Approach via Rank-Based Prototypes
Thiago César Castilho Almeida and Daniel Pedronette 33789

Do LLMs Cross Alphabets? A Cross-Script Audit of LLMs Across Four Kazakh Latin Proposals
Galammadin Askar, Arailym Mukhametkali, Akylbek Maxutov, Zhanat Makhataeva, Nūrali Medeu and Huseyin Atakan Varol 33804

Med-R2: An Adversarial Benchmark for Evidence-Grounded Reasoning in Medical VLMs
Wen MA, Fucheng Niu, Zhiting Fan, Zikai Xiao, Jiaxiang Liu and Zuozhu Liu 33820

Contaminated Collaboration: Measuring Gender Bias Transfer in LLM-Assisted Student Writing
Ariyan Hossain, Kazi Kamruzzaman Rabbi, Farig Sadeque and S M Taiabul Haque 33846

SlidesGen-Bench: Evaluating Slides Generation via Computational and Quantitative Metrics
Yunqiao Yang, Wenbo Li, Houxing Ren, Zimu Lu, Ke Wang, Zhiyuan Huang, Zhuofan Zong, Mingjie Zhan and Hongsheng Li 33865

Scroll-Net: Geometry-Aware Constrained Decoding for Multi-Page Document Hierarchy Parsing
Wanting Lin, Hiuyi Cheng, Wei Pan and Lianwen Jin 33908

On the Sensitivity of Instruction-tuned LLMs to Harmful Sentences in Long Inputs
Faeze Ghorbanpour and Alexander Fraser 33923

ChartREG++: Benchmarking and Improving Chart Referring Expression Grounding under Diverse referring clues and Multi-Target Referring
Tianhao Niu, Ziyu Han, Xuan Dong, Qingfu Zhu and Wanxiang Che 33945

Noise Adaptive Streaming Audio-Visual Speech Token Enhancement for Robust Full-Duplex Spoken Dialogue Models
Bella Godiva, Yeonju Kim and Yong Man Ro 33970

The Importance of Being Statistically Earnest: A Critical Re-evaluation of GSM-Symbolic
Dominika Agnieszka Długosz, Arlindo L. Oliveira and Natalia Díaz-Rodríguez 33984

Weaving Visual Narratives: Agentic Image Bundle Composition Beyond Atomic Visual Matching
Rong Shan, Tianyi Xu, Congmin Zheng, Wenteng Chen, Jiachen Zhu, Junjie Wu, Dun Zeng, Teng Wang, Weiwen Liu, Changwang Zhang, Weinan Zhang, Jun Wang and Jianghao Lin 34025

Modality-Specific Gating for Large Audio-Language Models
Zhenni Ren, Juan Wang, Qiuchi Li, Lei Li, Chunxiao Gao, Huaping Zhang and Baohua Zhang 34051

UniMem: Complementary Episodic-to-Parametric Memory for Boundary-Agnostic Task Streams
Siyu Xia, Chenheng Zhang, Yanting Wu, Haoxuan Li, Jiajun Chai, Xiaohan Wang, Guojun Yin, Wei Lin, Zhouchen Lin, Haifeng Zhang and Jun Wang 34066

When Does Hate Transfer? A Large-Scale Study of Cross-Lingual Hate Speech Detection
Zahra Bokaei, Walid Magdy and Bonnie Webber 34086

Multilingual Safety Signals Are Multi-Layered: Filtering Safety-Degrading Data for Safer LLMs
Jiakun Li, Guowei Song, Sijia Li, Xingwei He, Hongzheng Chai and Yuan Yuan 34105

E3-Rewrite: Learning to Rewrite SQL for Executability, Equivalence, and Efficiency
Dongjie Xu, Weijie Shi, Yue Cui, Qingzhi Ma, Wenwen Xia, Ruiyuan Zhang, Shimin Di, Jia Zhu, Kai Zheng and Jiajie Xu 34121

Beyond Static Grounding: Behavioral Contracts for Trustworthy Retrieval-Augmented Generation
Weiqing Zeng, Zida Yang, Guquan You, Haoran Yang and Yu Zheng 34136

Persona-CAPS: Controllable and Auditable LLM Personality via Retrieval over Situation–Behavior Contingencies
Guanghao Liu, Guangzhou An, Yang Xu and YU Chen 34153

An Image Is Worth Ten Thousand Words: Verbose-Text Induction Attacks on VLMs
Zhi Luo, Zenghui Yuan, Daizong Liu, Wenqi Wei and Pan Zhou 34185

OBJECTION! Lawyer Agents Mitigate Guilty Bias in Legal Judgment Prediction
Jaehoon Jeong and Jay-Yoon Lee 34205

CommunityBench: Benchmarking Community-Level Alignment across Diverse Groups and Tasks
Jiayu Lin and Zhongyu Wei 34226

When Discourse Pressures Conflict: Information Structure in Vision-Language Model Outputs
Marcell Fekete, Johannes Bjerva and Tamás Káldi 34248

How Identity and Opinion Shape Political Sycophancy in LLMs
Li-Ni Fu, Chang Chih Meng, Chien-Hua Chen, Hen-Hsen Huang and I-Chen Wu 34269

D3-Gym: Constructing Real-World Verifiable Environments for Data-Driven Discovery
Hanane Nour Moussa, Yifei Li, Zhuoyang Li, Yankai Yang, Cheng Tang, Tianshu Zhang, Nesreen K. Ahmed, Ali Payani, Ziru Chen and Huan Sun 34300

SPARROW: Scalable Taxonomy Induction via Structure-Preserving Partitioning and Constraint-Guided Merging
Yirui Zhang, Yixuan Tang, Yandong Sun, Mong-Li Lee and Anthony Kum Hoe Tung 34320

DFlare: Scaling Up Draft Capacity for Block Diffusion Speculative Decoding
Jiebin Zhang, Zhenghan Yu, Song Liu, Eugene J. Yu, Zheng Li, Dawei Zhu, Jiangshan Duo, Weimin Xiong, Yifan Song, Guanghua Yu, Jianchen Zhu and Sujian Li 34333

Subspace Alignment Score: Locating Domain-Relevant Important Parameters in Large Language Models
Le Wu, Jiayi Wu, Jixiang Xiong, Mengfa Mo and Jingmin Xin 34347

Ask Now, Use Later: Benchmarking the Proactivity Gap in Long-Lived LLM Agents
Bin Wu, Guanyun Zou, Bingbing Wang, Huan Zhao and Chuan Shi 34367

The Mirage of Calibrated Confidence: Trajectory-Independence of Verbalized Confidence in Vision-Language Models
Jisoo Yang, Jaeho Han, Trung X. Pham and Junyeong Kim 34404

INSPIRE: An Internalize-Then-Improve Approach for Example-Driven Mathematical Reasoning
Shuai Wang, Jiayi Kuang, Yinghui Li, Haojing Huang, Xinnian Liang, Ying Shen and Liang Lin 34424

DreamReasoner-8B: Block-Size Curriculum Learning for Diffusion Reasoning Models
Zirui Wu, Lin Zheng, Jiacheng Ye, Shansan Gong, Xueliang Zhao, Yansong Feng, Wei Bi and Lingpeng Kong 34440

Seeing the Unseen: Visual Similarity for Pixel Language Model Adaptation
Ran Zhang, Miryam de Lhoneux and Wessel Poelman 34454

Length and Meaning for Language Model Maps
Momose Oyama, Ryo Kishino, Hiroaki Yamagiwa and Hidetoshi Shimodaira 34472

Intrinsic Temporal Adaptation of CLIP for Partially Relevant Video Retrieval
Hyun Seok Seong, Woojin Jun, SuBeen Lee and Jae-Pil Heo 34488

Seeing is Believing: Benchmarking and Mitigating the Interference of Language Priors on Visual Perception in MLLMs
Li Ma, Yan Zhuang, Jiawen Deng and Fuji Ren 34500

Enhancing Long-Context Reasoning via Verifiable Atomic Skills
Yanling Xiao, Huaibing Xie, Guoliang Zhao, Shihan Dou, Nantao Zheng, Shaolei Wang, Yiting Liu, Cheng Zhang, Pluto Zhou, Zhisong Zhang and Lemao Liu 34516

Language Model Maps for Prompt-Response Distributions via Log-Likelihood Vectors
Momose Oyama, Yusuke Takase and Hidetoshi Shimodaira 34533

Tournament-GRPO: Group-Wise Tournament Rewards for Reinforcement Learning in Open-Ended Long-Form Generation
Zixuan Yang, Yiqun Chen, Wei Yang, Erhan Zhang, Zihan Shen, Xiaochi Wei, Yan Gao, Yi Wu, Yao Hu and Jiaxin Mao 34556

Stop Training Against Soft Users: Mentalized Adversarial Self Play for Proactive Dialogue
Zanwu Liu, Dan Zhang, Xiaodi Wang, Yongbo Mao, Xiaoyang Guo, Yonglin Wu, Peiyue Jiang, Yangyang Wei and Zecheng Lin 34574

Measuring Affinity between Attention-Head Weight Subspaces via the Projection Kernel
Hiroaki Yamagiwa, Yusuke Takase and Hidetoshi Shimodaira 34591

Inter-3D VQA: A Roadside Multimodal Benchmark for 3D Spatiotemporally Grounded Visual Question Answering
Shaozu Ding, Linan Song and Dajiang Suo 34661

SHICD: Self-Hallucination-Inducing Contrastive Decoding for Efficient LLM Hallucination Mitigation
Haopeng Jing, Shufeng Hao, Chongyang Shi and qi yu 34687

HachiWare: One-Shot Selective SVD in a Dual-Branch Framework for KV Cache Compression
Hsiao-Ting Shao, Xiaolin Lin, Yu-Chen Lu, Grace Li Zhang, Kai-Chiang Wu and Ning-Chi Huang 34701

One Instruction Is a Benchmark: End-to-End Instruction-Following Evaluation with FlexBench
Yuchen Ge, Yizhao Sun, Wei He, Jinpeng Ou, Wenzhe Niu, Chengshun Shi, Huichuan Fan, Biao Ma, Shuo Huang, Xin Tian, Jiuchong Gao, Jinghua Hao and Renqing He 34726

TRACE-GRPO: Teacher-Refined Advantage Credit Estimation for Long-Horizon Agentic RL
Yuchen Liu, Tianchu Yao, Deshu Chen, Yinghao Zhang, Xigui Li, Zehui Ling, Xin Guo, Yuan Cheng, Chao Qu, Lintao Ma and Yuan Qi 34746

VIABLE: A Visually Impaired Assistance Benchmark for VLM-as-a-Judge Evaluation
Yi Zhao, Siqi Wang, Zhe Hu, Yushi Li and Jing Li 34766

FAMPWQ: Fisher Information-based Adaptive Mixed Precision Weight Quantization for Effective LLM Inference
GongWei Lee, Ji Liu, Juncheng Jia and Ji Wu 34781

DuraS2ST: Chain-of-Thought and Reinforcement Learning for Duration-Aligned Speech-to-Speech Translation
Yayue Deng, Dingdong Wang, Yuxuan Hu, Jinyu Li, Yanqing Liu, Yuanyuan Wang, Weidong Chen, Helen M. Meng, Shujie Liu and Xixin Wu 34802

Fast-Slow Thinking RM: Efficient Integration of Scalar and Generative Reward Models
Jiayun Wu, Peng Zhang, Yuanyuan Lu, Shan Qu, Ning Gu and Tun Lu 34822

Representation-Aware Advantage Estimation: Your Reward Model Provides More Than A Scalar Output
Guozheng Li, Xiyan Fu and Steven Y. Guo 34839

SOLID: Infusing Symbolic Logic into LLMs via Dual-Verified Backward Reasoning
Weijie Yu, Yuhang Dong and Feifei Ma 34852

AgentEHR: Advancing Autonomous Clinical Decision-Making via Retrospective Summarization
Yusheng Liao, Chuan Xuan, Yutong Cai, Lina Yang, Zhe Chen, Yanfeng Wang and Yu Wang 34870

TVIR: Building Deep Research Agents Towards Text-Visual Interleaved Report Generation
Xinkai Ma, Zhiqi Bai, Dingling Zhang, Pei Liu, Yishuo Yuan, He Zhu, Jiakai Wang, Qianqian Xie, Yifan Zhao, Xinlong Yang, Hao Cong, Zhiheng Yao, Fengxia Xie, Zihao Xu, Haoran Xu, Zhaohui Wang, Minghao Liu, Shirong Lin, Yingshui Tan, Yuchi Xu, Wenbo Su, Zhaoxiang Zhang, Bo Zheng and Jiaheng Liu 34918

Generative Models Enhanced by Sequence Labelling and Aspect-Code Switching Improve Cross-lingual Aspect-Based Sentiment Analysis
Jakub Šmíd, Pavel Priban and Pavel Kral 34966

Manifold Embedding: A Point-to-Hyperplane Approach
Xianglong Shi, Yunhan Jiang, Nicu Sebe and Ziheng Chen 34982

TrustRAG: Metadata-Governed RAG for Compliance-Oriented Question Answering
Huiling Ge, Zhiguo Shi and Yungui Nie 35002

From Poetry Parties to Poetry Critics: Benchmarking Classical Chinese Poetry Generation and Evaluation
Zihao Yi, Zhaopeng Tu, Ruotian Ma, Qingxuan Jiang, Zhe Xu, Qingyu Xiao, Yixue Li, Ying Shen and Liefeng Bo 35021

Disentangling the Interpretive and Predictive Roles of LIWC: Controlled Substitution in Depression-Related Classification
Hsiang-Chen Yeh, Xiutian Zhao, Aurosweta Mahapatra, Shreeram Suresh Chandra, Ryan L. Boyd and Berrak Sisman 35037

Tampered ML: Evaluating LLM Judgment of Machine Learning Research
Soroush Ebadian, Parand A. Alamdari, Joe Benton and Hailey Joren 35058

SyRuP: Enhancing System-Prompt Following via Reward-Guided Prediction in LLM Decoding
Seoyeon Kim, Minjae Kang and Jaehyung Kim 35086

Frozen Cores Need Task Signal: Fisher-Whitened Cross-Covariance for Low-Resource LLM Adaptation
Wentao Ye, Zhanming Shen, Zhiqing Xiao, Yao Ding, Haobo Wang and Gang Chen 35114

Unlocking Multimodal Protein Language Models at Inference Time
Yi Zhou, Wang Qipeng, Yunqing Liu, Jun Xia, Li Qing and Wenqi Fan 35132

GSR-LoRA:Group-Sparse Routing-aware Low-Rank Adapter Compression for Multi-Task Mixture-of-Experts Fine-Tuning
Weimin Cheng, Chengyu Zhao, Xuanjing Huang and Yatian Shen 35152

From Global to Regional Coverage: Hyperspherical Prototype Discovery for Storage-Efficient Multimodal Retrieval
Xing Mu, Xiaotian Lin, Zhangyang Peng, Bairui Zhang and Yuyu Luo 35168

What and Whose Knowledge? Measuring Epistemic Diversity in Large Language Models
Dustin Wright, Sarah Masud, Jared Moore, Srishti Yadav, Maria Antoniak, Peter Ebert Christensen, Chan Young Park and Isabelle Augenstein 35186

When Compression Helps and When It Hurts: Condition-Aware Analysis of Chain-of-Thought Distillation
Siyang Lyu, Xinghao Chen, Zhijing Sun, Tong Liu, Dawei Zhu and Xiaoyu Shen 35204

Learning What to Share and What to Personalize: Hierarchical Strategy Co-Evolution for Agent Memory
Yupeng Han, Shuochen Liu, Kai Zhang, Ze Liu, Zhihong Pan and Xianquan Wang 35232

HEMP: Hierarchical Entity-aligned Memory Prompting for CT Report Generation
Tong Xu, Yiming Qian, Zhe Shan and Xia Xie 35250

HalloMTBench: A Large-Scale Human-Verified Benchmark for Diagnosing Hallucination Triggers in LLM Translation
Shaowei Zhang, Heng Liu, Xinwei Wu, Hao Wang, Yangyang Liu, Xiaohu Zhao, Jianxiang Peng, Tianhao Shen, Jiang Zhou, Bo Zeng, Tianyu Dong, Dan Shi, Linlong Xu, Longyue Wang, Deyi Xiong and Weihua Luo 35261

New Intent Discovery: From Local Relations to Global Structure
Peng Chen and Bang Wang 35282

Bayesian Control for Coding Agents
Theodore Papamarkou, Vladislav Smirnov, Viktor Mazanov, Artem Vazhentsev, Preslav Nakov, Timothy Baldwin and Artem Shelmanov 35299

Reactivating Alignment: Defending LLMs from Jailbreaks via Intention-Aware Input–Output Matching
Luoyu Chen, Weiqi Wang, Zhiyi Tian, Chenhan Zhang and Shui Yu 35342

Reconstructing the Right Episode: Evaluating Interleaved Conversational Memory Beyond Long Context
Zhexi Feng, Ruiyi Zhang, Yongbo Yang and Pengtao Xie 35367

SPRInG: Continual LLM Personalization via Selective Parametric Adaptation and Retrieval-Interpolated Generation
Seoyeon Kim and Jaehyung Kim 35386

Beyond the Illusion of Consensus: From Surface Heuristics to Knowledge-Grounded Evaluation in LLM-as-a-Judge
Mingyang Song and Mao Zheng 35416

The Persona-Switch Jailbreak Backdoor in LLMs: From Payload Retrieval to Trigger-Gated Persona-State Selection
Yanghao Su, Wenbo Zhou, Tianwei Zhang, Han Qiu, Weiming Zhang, Nenghai Yu and Jie Zhang 35434

MoralSteer: Safety-Preserving Fidelity Gains in Morally Complex Role-Playing via Theory-Guided On-Policy Self-Distillation
Shaojie Shi, Teqi Hao, Bohao Lv, Rui Xu, Zhengyu Shi and Xu Yinghui 35450

Seeing Before Synthesizing: VLM-Guided Transition Event Discovery for Weakly-Supervised Dense Video Captioning
Ye-Chan Kim, Seunghee Choi, SeungJu Cha, Si-Woo Kim, Hwiseon Kim, HyunGee Kim and Dong-Jin Kim 35464

ClueTracer: Question-to-Vision Clue Tracing for Training-Free Hallucination Suppression in Multimodal Reasoning
Gongli Xi, Kun Wang, Zeming Gao, Huahui Yi, Haolang Lu, Qianshan Wei, Ye Tian and Wendong Wang 35476

Structured Claim-Level Discourse Representations for Dense Health Narratives
Farnoush Nili, Elham Pourabbas Vafa, Shirin Nilizadeh and Eduard Dragut 35503

V-Rubrics: Visual Faithfulness via Rubric-Based Reinforcement Learning
Shulin Tian, Minglun Li, Yuhao Dong, Hao Ding, Jiarui Yao, Haiwen Diao, Jingkang Yang, Hongyuan Zhu and Ziwei Liu 35524

Tracing Error Propagation in Reasoning Language Models
Dung Huu Nguyen, Quan Pham and Dat Quoc Nguyen 35544

VTC-R1: Vision-Text Compression for Efficient Long-Context Reasoning
Yibo Wang, Yongcheng Jing, Shunyu Liu, Hao Guan, Rong-Cheng Tu, Chengyu Wang, Jun Huang and Dacheng Tao 35560

The Imitation Game: When LLMs Learn to Reason Like Programs via Code-Centric Reasoning Data Synthesis
Jinyang Zhang, Weibin Liao, Keqin Bao, Sihang Li, Shaobo Wang, Muyang Ye, Hongxin Ding, Yue Fang, Tianyi Tang, Fei Huang, Kexin Yang, Xingzhang Ren and Dayiheng Liu 35579

On the Role of Supporting Evidence in LLM Reasoning: A Causal Analysis of Legal QA with Activation Patching
Dalmo Buzato, Fernanda Sathler Rocha Franco, Caio Alves Caldeira, Isabela Carolina Ribeiro Santos, Evandro L. T. P. Cunha, Pedro Calais, Wagner Meira Jr. and Anisio Lacerda 35623

InfoSpec: Information-Guided Anchor Selection for Retrieval-Based Speculative Decoding
Jie Wang, Yi Ren, Wei Li, Zhenyang Gao and Jiawei Liao 35636

LatticeMind: A Conflict-Aware Memory Primitive for Multi-Agent Systems
Heng Zhou, Lian Zhang, Yutao Fan, Tiancheng He, Siki Chen, Hejia Geng, Philip Torr and Zhenfei Yin 35653

AdaRetouch: An Interleaved Vision-Language Agent for Adaptive Multi-Turn Image Retouching
Ziang Liu, Yongheng Zhang, Qi Wen, Kai Zhao, Jianwei Zhang, Yuezhou Dong, Yiren Hao, Bo Li, Peishan Dai and Hao Zhang 35673

AGENTQ: Quantization-Conditioned Backdoor Attacks on LLM Agents
Xiaoqun Liu and Qiben Yan 35693

UniReason-Med: A Shared Grounded Reasoning Interface for 2D-to-3D Transfer in Medical VQA
Mengzhuo Chen, Yan Shu, Chi Liu, Hongming Piao, Xidong Wang, Derek Li and Bryan Dai 35710

SimLens: Two-Token Continuation for Early Exit in Single-Token Decision Tasks
Ming Ma, Bowen Zheng, Zhongqiao Lin and Tianming Yang 35725

GENIE: A Fine-Grained Measure for Novelty
Ramya Namuduri, Manya Wadhwa, Anshun Asher Zheng, Greg Durrett and Junyi Jessy Li35742

Beyond Adapter Expansion: Packing Multiple Tasks into Spare LoRA Capacity
Shuangjie Fan, Chengxiang Huang, Zhiyuan Zhao, Piotr Koniusz, Junyu Gao and Di Hu . . 35766

Can Reasoning Models Detect Changes to their Chains of Thought?
Sathvik Manikantan Napa Ugandhar, Utkarsh Singh, Chengyuan Xue, Miriam Wanner and William Gantt Walden 35784

Cascade-SC: A Pareto-Efficient Cross-Model Cascade for Test-Time Reasoning
Minghan Li, Siqi Zhao, Luoliang Hua and Guodong Zhou 35802

Semantic Distractor Robustness in LLMs: Exploring the Underlying Mechanisms
Xiangyu Wong, Eugene J. Yu, Chenhao Xiong, Zheng Li, Yifan Song and Sujian Li 35820

Code-TVT: Synthesizing Test-Case-Driven Verifiable Tasks for Code Reinforcement Learning
Jin Jiang, Junlong Zhang, Yuchen Yan, Zhaoze, Changzhi Zhou, Suncong Zheng and Liangcai Gao 35842

TTT-Sparse: Test-Time Training of Sparse Attention for Long-Context Understanding
Xuehao Wang, Huayu Deng, Linhao Zhong, Yanbin Wei, Hengyu Zhao, Mingyu Liu, Hao Chen and Chunhua Shen 35866

Shopping Reasoning Bench: An Expert-Authored Benchmark for Multi-Turn Conversational Shopping Assistants
Shuxian Fan, Seonwoo Min, Youna Hu, Botao Xia, Jayakrishnan Unnikrishnan, Rowan Mussel-mann, Yifan Gao, Qingyu Yin, Priyanka Nigam and Bing Yin 35882

The Menu Is an Execution Prior: State-Path Tool Menus for Online Agents
Bo Yan, Weikai Lin and Song Wang 35904

PruneGround: Plug-and-play Spatial Pruning for 3D Visual Grounding
Duc Cao Dinh, Khai Le-Duc, Florent Draye, Chris Ngo, Terry Jingchen Zhang, Bernhard Schölkopf and Zhijing Jin 35918

Grammar Control for Language Learning with Sequential Monte Carlo and Small Language Models
Kate McCurdy, Jacob Hoover Vigly, Thomas Hikaru Clark, Karolina Stanczak, Yahya Emara, Joshua B. Tenenbaum, Alexander K. Lew, Ryan Cotterell and Timothy J. O'Donnell 35943

RetrievalRouter: Joint Modality and Architecture Selection for Document Retrieval
Emre Kuru, Mehmet Onur Keskin, Reza Farahbakhsh and Noel Crespi 35968

SIC-Agents: Benchmarking and Building Adaptive Simulator for Pediatric Serious Illness Communica-tion Training
Zihan Wang, Anita Slominska, Rennie Bimman, Elizabeth Di Flumeri, Amanda Mayappo-Neeposh, Conall Francoeur, Tamara Ellen Carver, Xiao-Wen Chang, Doina Precup, Esin Darici Haritaoglu, Ismail Haritaoglu, Akshatha Arodi and Naomi Goloff 35983

GRACE: Step-Level Benchmark for Faithful Reasoning over Context
Hoang Pham, Dong Le and Anh Tuan Luu 36002

Understanding the Limits of Agentic ICD Coding
Eng Chong Yock, Yushi Cao, Yiming Chen, Kezhi Mao and Hongchao Jiang 36028

MELD: Mel-Spectrogram-Based Speech Language Modeling with Discrete Latent Variables
Sung-Lin Yeh, Wei Zhou, Gil Keren, Duc Le, Zhong Meng, Hao Tang, Jay Mahadeokar, Ozlem Kalinli and Alexandre Mourachko 36038

When Queries Crowd: Diagnosing and Improving Spread-Out Regularization for Text Embeddings
Zeqi Zhou, Qingcheng Zeng, Rui Yang, Jiarui Liu, Heli Qi, Puxuan Yu, Weihao Xuan and Naoto Yokoya 36051

Beyond the Stability-Exploration Dilemma: Environmental Regularization for LLM Policy Optimization
Xianlei Zhou, Xiangdi Meng, Yu He, Tianyu Qi, Shuyan Guan, Xianli Zhang, Jian Zhang, Xin Li, Qika Lin and Jun Liu 36071

Beyond the Current Observation: Evaluating Multimodal Large Language Models in Controllable Non-Markov Games
Shengyuan Ding, Xilin Wei, Xinyu Fang, Haodong Duan, Dahua Lin, Jiaqi Wang and Yuhang Zang 36087

Surfacing the Unsaid: CUE-Bench for Affective Stance in Chinese Discourse
Zhenyan Zheng, Yunyao Zhang, Junxi Sheng, Junqing Yu and Zikai Song 36109

Every Answer Has Its Own Path: Agentic Routing for Retrieval-Augmented Table Question Answering
A Young Kim, Jisu Shin, Donghee Han and Mun Yong Yi 36127

Identifying and Mitigating Bottlenecks in Role-Playing Agents: A Systematic Study of Disentangling Character Profile Axes
Yonghyun Jun, Junhyuk Choi, Jeonghyun Park, Jihyeong Park, Nicole Geumheon Liu and Hwanhee Lee 36152

ReflectRec: Cycle-Consistency Preference Verification for LLM-based Recommendation
Dae Kyoo Kim, Heejung Cho, Sungmoo Hong and Eunil Park 36183

Bag-of-Tokens Tuning: Calibrating LLM Logits with Contextual Probabilities
Chen-Meng Qiu, Ze-Hao Wang, Deng-Bao Wang, Linan Yue and Min-Ling Zhang 36198

ACE-Align: Attribute Causal Effect Alignment for Cultural Values under Varying Persona Granularities
JiaTang Luo, Bingbing Xu, Rongxin Chen, Xiaoyan Zhao, Yang Zhang, Liang Pang, Zhiyong Huang and Huawei Shen 36216

Mindscape-Aware Retrieval Augmented Generation for Improved Long Context Understanding
Yuqing Li, Jiangnan Li, Zheng Lin, Ziyan Zhou, Junjie Wu, Jie Zhou and Mo Yu 36236

How AI Models Manage Epistemic Authority: A Taxonomy and Comparative Analysis of Responses to User Disagreement
Riyadh Alnasser, Yusuf Mücahit Çetinkaya, Sumin Zhao and Tugrulcan Elmas 36257

FBHM: Functional Benchmarking and Steering of VLMs for Hateful Meme Detection
Paramananda Bhaskar, Naquee Rizwan, Daksh Jogchand, Saurabh Kumar Pandey and Animesh Mukherjee 36273

Perceive, Localize, Reflect: Evidence-Anchored Multi-Agent Debate for MLLM Hallucination Mitigation
Yebo Wu, Han Jin, Wei Yang Bryan Lim and Li Li 36298

Send a SCOUT First: Pre-hoc Reasoning for Adaptive Detector Allocation in Prompt-Injection Defense
Shuhao Zhang, Jiarui Li, Qi Cao, Ruiyi Zhang and Pengtao Xie 36314

Prompt Revision as a Source of Cultural Bias in Text-to-Image Systems
Aleksandra Urman, Elsa Lichtenegger, Salima Jaoua, Azza Bouleimen, Robin Forsberg, Corinna Hertweck, Stefania Ionescu, Nicolò Pagan, Ancsa Hannak and Joachim Baumann 36352

Personalized Alignment of Language Models with Human Value Profiles
Rashon Poole and Farnaz Jahanbakhsh 36378

ChJGBench: A Comprehensive Benchmark for Chinese Legal Judgment Generation with Diagnostic Evaluation
Yinghao Hu, Linggang Kong, Qian Xiao, Yuting Huang, Yunlong Zhang, Yujie Zhang, Yiquan Wu, Fei Duan, Kaisong Song, Jun Lin, Kun Kuang and Fei Wu 36404

SpiralFormer: Looped Transformers Can Learn Hierarchical Dependencies via Multi-Resolution Recursion
Chengting Yu, Xiaobo Shu, Yadao Wang, Yizhen Zhang, Haoyi Wu, You Wu, Rujiao Long, Ziheng Chen, Yuchi Xu, Wenbo Su and Bo Zheng 36429

Knowledge Editing for Masked Diffusion Language Models
Haewon Park and Yohan Jo 36452

CORE-Bench: A Comprehensive Benchmark for Code Retrieval in the Era of Agentic Coding
Fuwei Zhang, Yanzhao Zhang, Mingxin Li, Dingkun Long, Lexiang Hu, Pengjun Xie, Fuzhen Zhuang and Zhao Zhang 36477

Safety Paradox: How Enhanced Safety Awareness Leaves LLMs Vulnerable to Posterior Attack
Long P. Hoang, Viet Hai Le, Shaoyang Xu, Wei Lu and Wenxuan Zhang 36506

A Mechanistic Study of AI-Text Detection Neurons in Frozen BERT: Sparse Probing and Activation Patching on RAID
Paweł Blicharz and Miłosz Grunwald 36533

Query-focused and Memory-aware Reranker for Long Context Processing
Yuqing Li, Jiangnan Li, Mo Yu, Yanyu Chen, Guoxuan Ding, Zheng Lin, Wei Zhang and Jie Zhou 36553

Stride-k Subsampling: Train-Free Audio Token Reduction for Whisper
Chanhee Cho, Junhyuk Choi and Bugeun Kim 36572

The Attentional White Bear: Persistent But Attenuated Concept Recoverability in Transformer Language Models
Rebecca Ramnauth and Brian Scassellati 36600

TRACE: A Unified Rollout Budget Allocation Framework for Efficient Agentic Reinforcement Learning
Heming Zou, Qi Wang, Yun Qu, Yuhang Jiang, Lizhou Cai, Yixiu Mao, Ru Peng, Xin Xu, Weijie Liu, Kai Yang, Saiyong Yang and Xiangyang Ji 36615

UniMER: Multilingual Multimodal Emotion Recognition with Prototype-Guided Attention Pooling
Yinghua Qin and David C. Anastasiu 36646

The Cost of Sounding Different: Accent Bias in Audio Language Models
Aruna Srivastava, Orevaoghene Ahia, Li Lucy, Ashley Christendat, Sameep Chattopadhyay, Farhan Samir, Tejumade Afonja, Valentin Hofmann, Sachin Kumar, Noah A. Smith and Yulia Tsvetkov 36660

SPIRE: Unifying Partial Rollouts, Replay, and Asynchrony in LLM Reinforcement Learning via Segment-wise Prefix Importance Sampling
Chenlu Zhan and Yufei Zhang 36688

The Shape of Time: Video-Token Contrast for Temporal Understanding in VideoLMs
Yumeng Shi, Quanyu Long, Yin Wu and Wenya Wang 36702

Exploring Confirmation Bias: LLMs Perceive Democratic Claims as More Truthful Independently of their Content
Charlott Jakob, Jing Yang, Daniel Renken, Veronika Solopova, Max Upravitelev, Sebastian Möller and Vera Schmitt 36718

A Multimodal Automatic Redteaming Evaluation based on Atomic Jailbreak Strategy Decoupling and Combination
Shiji Zhao, Yuxuan Zhou, Chen Xiong, Dongxian Wu, Yang Bai and Xun Chen 36733

QAQ: Bidirectional Semantic Coherence for Selecting High-Quality Synthetic Code Instructions
Jiayin Lei, Ming Ma, Yunxi Duan, Chenxi Li and Tianming Yang 36749

He Said, She Said, They Simulated: Simulating Situated Gossip with LLM Agents
Yicong Tan, Yun Shen and Yang Zhang 36763

FaithMed: Training LLMs For Faithful Evidence-Based Medical Reasoning
Zhiyun Zhang, Liwen Sun, Xiang Qian and Chenyan Xiong 36807

APPSim-Bench: Bridging Real-world Apps and Reproducible Evaluation for Mobile GUI Agents
Jintian Feng, Long Chen, Xiao Yu, Jiayi Dai, Chenglong Liu, Haoru Wang, Zizhen Xue, Yuxuan Shi, Ziyang Wang and Yichen Gong 36823

Is Convergence Inevitable? Tracing Output Homogeneity Back to Base Models
Alexandrine Fortier, Hazel Chen and Peter West 36846

Users Mispredict What Drives Their Preferences for AI Writing Assistance
Vivian Lai, Zana Buçinca, Hyeonsu B Kang, Mo Houtti, Nil-Jana Akpinar, Kevin Chian, Namjoon Suh and Alex C. Williams 36866

Bias in the Tails: How Name-conditioned Evaluative Framing in Resume Summaries Destabilizes LLM-based Hiring
Huy Nghiem, Phuong-Anh Nguyen-Le, Sy-Tuyen Ho and Hal Daumé III 36889

SAND: Subjective-Anticipation-Augmented News Recommendation via Dual-Pathway Distillation
Hong-Kyun Bae, Zhenping Jin and Sang-Wook Kim 36934

DERELAB: Probing Defeasible Reasoning and Confirmation Bias in LLMs with a Generative Benchmark
Jayanta Sadhu, Sayem Shahad and Kenneth Marino 36952

The Hallucination Signal Is a Mean Shift: Why Simple Probes Suffice
Jungseob Lee, Jaehyung Seo and Heuiseok Lim 36986

DEEPCHART: How Far are LLMs from Faithful Data-Science Chart Generation?
Jiahui Tang, Kuicai Dong, Dexun Li, Hongchao Gu, Haocheng Yu, Wei Han, Chen Zhang, Yong Liu, Hao Wang and Enhong Chen 37005

Do LLMs Trust the Accuser or the Accusation? Measuring Belief Shifts in Werewolf
Yu-Yu Yang, Ti-Rong Wu, Hung Guei, Hsing-Yu Chen and I-Chen Wu 37032

Beyond Scores: Understanding LLM-as-a-Judge Mechanisms in Summarization Evaluation
Himil Vasava and Ming Jiang 37044

CharLOTTE: Orthographic Correspondence-Based Data Unification Enhances Cross-Lingual Transfer in Low-Resource NMT
Brendan T. Hatch, Peter K. Bickel, Eric K. Ringger and Stephen D. Richardson 37061

iPOE: Interpretable Prompt Optimization via Explanations
Jiahui Li, Yarik Menchaca Resendiz, Sean Papay and Roman Klinger 37087

Typological Alignment of Stack-Based Language Models on Mildly Context-Sensitive Artificial Languages
Nadine El-Naggar, Tatsuki Kuribayashi and Ted Briscoe 37110

Tracing Computation Density in LLMs
Corentin Kervadec, Iuliia Lysova, Iuri Macocco, Marco Baroni and Gemma Boleda 37124

Beyond Planning: Sparse Context Reset for Long-Horizon GUI Agents
Weiyao Huang, Xinye Wu, Zeming Liu, Xiaoyu Shen and Yunpu Ma 37145

ContextRouter: Learning When to Use Context for Efficient LLM Machine Translation
Ritvik Choudhary, Rem Hida and Toshiyuki Sekiya 37164

Do Vision Language Models Understand Human Engagement in Games?
Ziyi Wang, Qizan Guo, Rishitosh Kumar Singh and Xiyang Hu 37184

AgentTAC: Mining Temporal Action-Context Error Patterns from Agent Traces
Chenchen Zhao, Muxi Chen, Jianhai Shu, Zhaohua Zhang, Wenyu Jiang, Tianwen Jiang, Hexuan Deng, Sa Zhu, Haolin Yang, Geyuan Zhu, Cheng Yang, Qiuyong Xiao, Jianhuan Zhuo, Jihong Zhang and Qiang Xu 37202

PERIBENCH: A Protocol-Aware Benchmark for Dynamic Perioperative Risk Prediction from Numerical Models to Language Interfaces
Xiaohui Zhao, Xiangyu Meng, Jinjin Chi, Yongcheng Jing, Yiming Wang, Yibing Zhan, Hang Sun and Dacheng Tao 37223

UTILMEM: Benchmarking Evidence Utilization in Long-Term Conversational Memory
Peijun Qing, Fobo Shi and Soroush Vosoughi 37249

Is the Future Predictable? Modeling Future Predictability for Reliable Temporal Reasoning
Yongyi Huang, Huaping Zhang, Wenyao Cui, Qiuchi Li, Lei Li and Baohua Zhang 37273

Osprey: Target-agnostic Pre-training Makes Stronger Drafters in Speculative Decoding
Fengxiang Bie, Yuqing Jian, Yifan Yu, Zhongzhu Zhou, Zelei Shao, Ben Athiwaratkun, Shuaiwen Leon Song, Chenfeng Xu, Xiaoxia Wu and Tianyi Zhang 37292

Forget Without Compromise: Nexus Sampling for Streaming KV-Cache Eviction Under Fixed Budgets
Duc Duong, Hoang Anh Duy Le, Jianwen Xie, Anshumali Shrivastava and Zhaozhuo Xu . 37313

From Specialization to Generalization: Instruction-tuned LLMs for Robust Harmful Content Mitigation
Lukas Edman, Daryna Dementieva and Alexander Fraser 37335

World Models Meet Language Models: On the Complementarity of Concrete and Abstract Reasoning
Yucheng Zhou, Wei Tao, Steven Y. Guo and Jianbing Shen 37356

Disentangling Curriculum Learning in NLP: Towards a Unifying Taxonomy
Vanessa Toborek, Florian Seiffarth, Sebastian Müller and Tamas Horvath 37384

What's in a Name? Morphological Shortcuts by LLMs in Pharmacology
Kaijie Mo, Thomas Yang, Chantal Shaib, Qing Yao, William Rudman, Ramez Kouzy, Kanishka Misra, Byron C Wallace and Junyi Jessy Li 37411

Counter with Evidence! A Multi-Agent Memory Efficient Reasoning Framework for Hate Category Informed Counterspeech Generation
Sujoy Nath, Aswini Kumar and Tanmoy Chakraborty 37435

CPJudgeBench: Meta-Evaluation of LLM Judges for Constraint Programming through Solution-Space Semantics
Nastaran Moradzadeh Farid, Alireza Taghizadeh, Tommaso Di Noia, Yashar Deldjoo and Sara Shafiee 37459

Adaptive Boundaries: Context-Aware Detection for Synthetic Text
Haorong Li, Ruitao Ni, Haihui Yang, Shiping Gao, Yang Liu and Xiaojun Quan 37479

Lost in Delusion: Examining LLM Safety Under User Delusions and Distress
Andrew Aquilina, Chetna Nihalani, Vasudha Varadarajan, Nathan S. Fishbein, Yu-Ru Lin and Maarten Sap 37499

Large Reasoning Models Learn Better Alignment from Flawed Thinking
ShengYun Peng, Pin-Yu Chen, Eric Michael Smith, Song Jiang, Hongyuan Zhan, Haozhu Wang, Mahesh Pasupuleti, Duen Horng Chau and Jianfeng Chi 37538

ProactiveEngine: Proactive Learning with Lifelong Memory via World Models
Python Song and Xiaoying Liao 37567

Evaluating Coding Agents on Kernel Exploit Generation
Junyoung Jang, Gwanhyun Lee, Hwiwon Lee, Kyuheon Kim, Jongseong Kim, Jinho Jung and Lingming Zhang 37581

Fully Decoupled Reinforcement Learning with Adaptive Data Curation for GUI Agents
Pengxiang Li, Zechenhu, Zirui Shang, Jingrong Wu, Yang Liu, Hui Liu, Zhi Gao, Chenrui Shi, Bofei Zhang, Zihao Zhang, Xiaochuan Shi, Zedong YU, Yuwei Wu, Xinxiao Wu, Yunde Jia, Liuyu Xiang, Zhaofeng He and Qing Li 37604

Can Crowdsourcing Survive the LLM Era? A Community Survey on Human Data Collection
Aswathy Velutharambath, Neele Falk, Sofie Labat, Tarun Tater and Amelie Wuehrl 37634

DuQUAD: A Dual-View Framework with Quality-Aware Evidence Pruning for Multi-Document Question Answering
Yeonjun Choi, Seyeon Jeong, Jong Wook Kim and Beakcheol Jang 37655

Are Online Skill and Memory Modules Always Worth Their Tokens? A Budget-Constrained Study of Web Agents
Sina Hajimiri, Masih Aminbeidokhti, Jose Dolz, Ismail Ben Ayed, Issam H. Laradji, Spandana Gella and Nicolas Gontier 37676

R-IDEAL: Reinforcing Insight-Driven Experience for Agent Learning
Wentao Hong, Xin Zhou, Yukai Xu, Jianing Yu, Sheng Yang, Sheng Gao and Weiran Xu . 37694

Quantifying and Mitigating Korean Jamo-Level Typographical Vulnerabilities in Large Language Models
Seojin Lee and Hwanhee Lee 37709

Smooth Operator: Smooth Verifiable Reward Activates Spatial Reasoning Ability of Vision-Language Model
Jiao Siwen, Tianxiong Lv, Kangan Qian, Chenxu Zhao, Xiuyuan Zhu, Yangyi Fang, Jinyu Li, Zhihao Liao and Yang Cai 37727

Edits in the Wild: A Multilingual Benchmark for Localizing On-Device Generative Image Edits on Consumer Platforms
Mohd. Farhan Israk Soumik, Awal Ahmed Fime, Malithi Mithsara Wanniarachchi Kankanamge, Arun Babu Aryal, Hoda Joudi Haghighi, Bibek Koirala, Ahmed Imteaj and Abdur R. Shahid . . . 37756

Mitigating Hallucination in Vision-Language Models through Barrier-Regulated Adaptive Closed-form Steering
Soumyadeep Jana, Pulkit Mittal and Ranbir Singh Sanasam 37790

Metonymic Circuits for Abstract Concept Grounding in Vision Transformers
Jing Ding, Ziqiao Ma, Jiayuan Mao, Joyce Chai and Freda Shi 37802

Proof-of-Use: Mitigating tool-call hacking in Deep Research Agents
Shengjie Ma, Chenlong Deng, Jiaxin Mao, Jiadeng Huang, Xuhui Jiang, Teng Wang, Junjie Wu, Changwang Zhang and Jun Wang 37843

DECENT: Distribution-aware Prototypical Calibration with Fine-grained Voting for VLM Misclassification Detection
Yijun Gong, Bohan Yang, Zhi Zhang, Yupeng Qin, Wenpeng Xing, Changting Lin, Meng Han, Ziyue Qiao and Xiao Luo 37858

SeVeR: Selective Visual Exposure and Retrieval for 3D Medical Image Question Answering
Yaojun Hu, Danyang Tu, Yang Liu, Jiajin Zhang, Wei Fang, Zhiqiang Liu, Chunlai Dong, Yingda Xia, Haochao Ying, Jian Wu and Ling Zhang 37874

PRO-Step: Step-level Process Reward Optimization for Retrieval-Augmented Generation
MinKeon Kim, Namjun Lee and Jaekwang Kim 37904

The Illusion of Expertise: Narrow Finetuning Induces Persistent Domain Priors in Vision-Language Models
Qinglin Zeng, Jusheng Zhang, Jiangqi Huang, Ningyuan Liu, Luozhang Xu, Wenhao Wang and Keze Wang 37926

WebCompass: Evaluating Multimodal Code Agents in Web Development from Vision, Instruction, and Interaction to Implementation
Xinping Lei, Xinyu Che, Junqi Xiong, Chenchen Zhang, Yukai Huang, Chenyu Zhou, Haoyang Huang, Minghao Liu, Letian Zhu, Hongyi Ye, Ken Deng, Zizheng Zhan, Han Li, Yifan Yao, Ming Sun, Zhaoxiang Zhang and Jiaheng Liu 37948

Small RL Controller, Large Language Model: RL-Guided Adaptive Sampling for Test-Time Scaling
Runpeng Dai, Tong Zheng, Rui Liu, Chengsong Huang and Hongtu Zhu 37985

Hybrid Open-Ended Tri-Evolution Makes Better Deep Researcher
Hongming Piao, Chi Liu, Mengzhuo Chen, Yan Shu, Xidong Wang, Derek Li, Ying Wei and Bryan Dai 38000

KnowledgeGain: Evaluating and Optimizing Science News Generation for Reader Learning
Dominik Soós, Meng Jiang and Jian Wu 38035

Not all Negation Cues are Equal: Affixal Negations Yield Better Negation Understanding
Tian Tan and Eduardo Blanco 38056

From Value Conditioning to Behavioral Shift: Lightweight Value Alignment of LLMs
Shangrui Nie, Florian Mai, David Kaczér, Charles Welch, Zhixue Zhao and Lucie Flek . . . 38077

TagPR: Tag-Guided Process Supervision for Personalization Reasoning in Large Language Models
Song Jin, Juntian Zhang, Ruyu Lyu, Yong Liu, Xun Zhang, Yufei Zhang, Fei Jiang, Guojun Yin, Wei Lin and Rui Yan 38104

Detecting AI Impostors: How Do Middle Schoolers Identify LLM Agents in a Live Collaborative Setting?
Dan Schumacher, Pragathi Durga Rajarajan, Haven Kotara, Roman Rendon, Kosi Atupulazi, Deepti Tagare, Ismaila Temitayo Sanusi, Fred G. Martin and Anthony Rios 38128

Perturbation: A simple and efficient adversarial tracer for representation learning in language models
Joshua Rozner and Cory Shain 38151

Trustworthy Medical LLMs via Bayesian Belief Updating
Mengyao Guo, Shufang Xie, Xiuying Chen, Shuqi Li, Hao Jiang, Bo Du and Rui Yan 38183

The Grammar of Transformers: A Systematic Review of Interpretability Research on Syntactic Knowledge in Language Models
Nora Graichen, Iria de-Dios-Flores and Gemma Boleda 38199

RASST: Retrieval-Augmented Simultaneous Speech Translation
Jiaxuan Luo, Siqi Ouyang, Jiaxing Xu and Lei Li 38241

Who is the Agent to Blame? Localizing Faithfulness and Citation Mistakes in Agentic Deep Research
Eran Hirsch, David Wan, Han Wang, Elias Stengel-Eskin, Mohit Bansal and Ido Dagan . . . 38261

DenseMixer: Improving MoE Post-Training via Precise Router Gradient
Feng Yao, Junxia Cui, Ruohan Zhang, Liyuan Liu, Shibo Hao, Li Zhang, Chengyu Dong, Churan Zhi, Shuohang Wang, Yelong Shen, Jianfeng Gao and Jingbo Shang 38287

UXBench: Benchmarking User Experience in AI Assistants
Mengze Hong, Xia Zeng, Zeyang Lei, Sheng Wang, Chen Jason Zhang, Di Jiang, Taiming Fu, Jinfeng Huang, Mengqiao Liu, Qinghe Chang, Haosheng Zou, Qiongyi Zhou, Sijun He, Xiaoshuai Chen, Minlong Peng and Di Liang 38299

RECAP: Regression Evaluation for Continual Adaptation of Prompts
Harsh Deshpande, Kushal Chawla, Sangwoo Cho, William M. Campbell and Sambit Sahu 38317

Self-Evolving Vision Language Models for Structured Visual Understanding
Zongxia Li, Xiaomin Wu, Chengsong Huang, Hongyang Du, Xiyang Wu, Lantao Yu, Yicheng He, Zhichao Liu, Jiarui Zhang, Fuxiao Liu and Jing Xie 38332

TranslatePsy-AfriSLM: High-Quality Data Scaling For Low-Resource Machine Translation
Milan Gritta, Patrik Lambert, Jihye Back and Amril Nazir 38348

Same Trajectory, Contradictory Rewards (RoboRMBench): Paraphrase Fragility in Vision Language Reward Models
Wonje Jeung, Sangyeon Yoon, Hyesoo Hong, Yoonjun Cho, Dongjae Jeon, Bumjun Kim, Jean Oh, Youngjae Yu and Albert No 38376

DSR-SQL: Dual-Schema Reasoning for Text-to-SQL and PI-MCE Optimized Schema Linking
Hu Lu, Fanghao Li, Dongqi Han and Yao Zhang 38396

Consensus Drift: Semantic Salience Can Override Majority Evidence in LLM Aggregation
Gokul Srinath Seetha Ram, Rashmi Elavazhagan and Anirudh Vasudevan 38411

SlideForge: Controllable Editing of Slides as Structured Artifacts
Haozhen Zheng, Fulin Wang, Tianhu Xiong, YingJie Yu, Shengyi Qian, Hanchao Yu, Alex Schwing, Klara Nahrstedt and Mingyuan Wu 38429

DR^3-Eval: Towards Realistic and Reproducible Deep Research Evaluation
Qianqian Xie, He Zhu, Xueming Han, Qingheng Xiong, Tiantian Xia, Fanyu Meng, Jiakai Wang, Zhiqi Bai, Chengkang Jiang, Zhaohui Wang, Yubin Guo, Yuqing Wen, Jiayang Mao, Zijie Zhang, Shihao Li, Yanghai Wang, Yuxiang Ren, Junlan Feng and Jiaheng Liu 38472

Not Just the Destination, But the Journey: Reasoning Traces Causally Shape Generalization Behaviors
Pengcheng Wen, Yanxu Zhu, Jiapeng Sun, Han Zhu, Yujin Zhou, Chi-Min Chan, Yike Guo and Sirui Han 38505

Elastic Horizon: Discovering the Effective Interaction Frontier in Agentic Reinforcement Learning
Gangyi Zhang, Junjie Meng, Letian Zhang, Wei Wu, Yang Zheng, Dong Wang, Yang Liu, Guanjun Jiang and Chongming Gao 38533

Less is MoE: Trimming Experts in Domain-Specialist Language Models
Haoze He, Xinkai Zou, Xuan Jiang, Xingyuan Ding, Ao Qu, Juncheng B Li and Heather Miller 38550

Cross-lingual Representation Learning via Centroid Intervention Fusion
Wei Sun and Marie-Francine Moens 38583

TRIAGE: Label-Free Hateful Video Detection with Boundary Mapping and Resolution
Junyi Chen, Qian Liu, Jing Sun and Yi Zhang 38600

How Do People Challenge Racial Stereotypes Online? Counter-Story Detection Across Reddit Communities
Uma Gunturi, Jimin Mun, Maarten Sap and Maria Antoniak 38622

Can Segmentation Models Understand the World? Proactive Affordance Reasoning via Vision-Grounded Chain-of-Thought
Yuchen Guo, Junli Gong, Hongmin Cai, Yiu-ming Cheung and Weifeng Su 38650

Mitigating Memory Interference in Lifelong Model Editing through Multi-Head Hopfield Retrieval
Yue Che, Chen Guoyi, Xijian Deng, Yujie Wu, Jiang Cai, Xuming Ran and Mingkun Xu . . 38663

OrgAccess: A Benchmark for Role-Based Access Control in Organization Scale LLMs
Debdeep Sanyal, Umakanta Maharana, Hong Ming Tan, Yash Sinha, Shirish Karande, Mohan Kankanhalli, Dhruv Kumar and Murari Mandal 38682

Cast a Wider Net: Coordinated Pass@K Policy Optimization for Code Reasoning
Yilong Li, Suman Banerjee and Tong Che 38736

How Well Do Vision-Language Models Understand Movies?
Emmanouil Zaranis, António Farinhas, Saul Santos, Beatriz Canaverde, Miguel Moura Ramos, Wafaa Mohammed, Giuseppe Attanasio, Chrysoula Zerva, Nithin Sivakumaran, Shoubin Yu, Elena Bueno-Benito, Aditya K Surikuchi, Ben Peters, Danae Sanchez Villegas, Andre G. Viveiros, Pavlo Vasylenko, Baohao Liao, Sonal Sannigrahi, Jaehong Yoon, Elias Stengel-Eskin, Mariella Dimiccoli, Oswald Lanz, Alessandro Suglia, Mohit Bansal, Sandro Pezzelle, Stella Frank, Vlad Niculae, Desmond Elliott, Raffaella Bernardi, Raquel Fernández and Andre Martins 38762

Beyond Fresh Starts: Stateful Inference for Streaming ASR in Conversational Voice Agents
Sameep Chattopadhyay, Alexander Erdmann and Mari Ostendorf 38802

TRACE: Triplet Recovery via Arrowhead-Centric Extraction for Flowchart Understanding
Daozhu Dong, Kaiwen Shi, Dan Si, Xiaoyu Hao, Tong Liu, Wenjie Zhang and Gong Cheng 38812

HyQuant: Hybrid-Precision Quantization for LLM Attention
Jiatong Ding, Bingxin Xing, Yu Zhang, Dian Ding, Xiaodong Yi, Xianbin Ouyang, Fei Hu Zhou, Kun Zhang, Zhenyu Guo, Hao Pan, Guangtao Xue and Yiming Zhang 38834

Format Collapse in Policy Optimization for LLM Reasoning
Debdeep Sanyal, Aakash Sen Sharma, Saurabh Deshpande, Dhruv Kumar and Murari Mandal 38849

AdiDhwani: Preservation-Oriented ASR for Endangered Bhili and Mundari Languages
Pooja Singh, M Kaab Bin Shahid and Sandeep Kumar 38870

TCD: Empowering Zero-Shot Detection of LLM-Generated Text using Token Contextual Dependency
Shihan Ma and Quan Wang 38911

Learning Concepts, Not Tokens: Self-Supervised Semantic Alignment for Language Models
Christine Zhang, Dan Jurafsky and Chen Shani 38927

Towards Understanding Pause Token Fine-Tuning Dynamics: A Mode Retention Perspective
Jaehyeon Kim, Suhwan Kim, Nakyung Lee, Yeongoon Kim, Jimin Seo, Giho Lee and Jungwoo Lee 38957

Beyond Seeing: Evaluating Multimodal LLMs on Tool-Enabled Image Perception, Transformation, and Reasoning
Xingang Guo, Utkarsh Tyagi, Advait Gosai, Paula Vergara, Jayeon Park, Ernesto Gabriel Hernández Montoya, Chen Bo Calvin Zhang, Bin Hu, Yunzhong He, Bing Liu and Rakshith Sharma Srinivasa 38981

LTS-VoiceAgent: A Listen-Think-Speak Framework for Efficient Streaming Voice Interaction via Semantic Triggering and Incremental Reasoning
Wenhao Zou, Yuwei Miao, Zhanyu Ma, Jun Xu, Jiuchong Gao, Renqing He and Jingwen Xu 39049

Repetition Mismatch: Why Data Mixture Experiments Don't Scale and How to Fix Them
Kevin Zhou, Lisa Alazraki, Kris Cao and Marek Rei 39070

Less Uniform Discrete Diffusion is More Powerful and Scalable
Kaibo Wang, Ding Ding, Fangyu Ding, Zijin Feng, Han Shi, Haoli Bai, Jiacheng Sun and Yang Xiang 39089

Script Is Not Preprocessing: Tokenizer-Mediated Transfer in Endangered Language Machine Translation
Pooja Singh, Saurabh Chauhan, Nishant Agnihotri and Sandeep Kumar 39110

Plans You Can Check: Verifier-Grounded Learning of an Open-Weight Planner for Executable Video-Editing
Haoyu Wang, Cheng Feng, Liuyang Bian, Ruiyang Huang, Lei Wei, Yafei Wen, Xiaoxin Chen and Xiaoying Tang 39151

Reliability of LLM Judges for Evaluating Entity Alignment
Vaibhava Lakshmi Ravideshik and Mayank Kejriwal 39175

Does Sensitivity to Pragmatic Norms Emerge with More Parameters, More Data, or More Conversations?
Raha Askari, Judith Sieker, Valerio Basile and Sina Zarrieß 39201

PRISM Edit: One Vector for All Temporal Answers
Chen Huang, Qi Zheng, Zhengruiqin, Long Zeng and Yuantong Xu 39217

Trust Before Fusion: QIMG-7 and Source-Aware Resolution for Polluted Multimodal RAG
Saadeldine Eletter, Owais Aijaz and Preslav Nakov 39236

Energy-Based Transformers as Predictors of Reading Difficulty
Jakub Dotlacil and Ece Takmaz 39262

LLMs Learn Better In-Context from Rules than from Examples
Xiang Fu, Seungmin Cho, Yukyung Lee and Najoung Kim 39279

SALT: Salience-Aware Lexical Trie for Long-Context Compression
Oteo Mamo, Hyunjin Yi, Joydhriti Choudhury, Shangqian Gao and Weikuan Yu 39314

Apples to Apples? Towards Comparable Crosslingual Language Model Evaluation
Xiulin Yang, Ethan Gotlieb Wilcox and Catherine Arnett 39332

AI Writers Have a Consistent Stylometric Footprint, but AI Editors Do Not
Zhengyang Shan, Yukyung Lee and Sophie Hao 39352

Discovering Hidden Gems in Model Repositories
Jonathan Kahana, Eliahu Horwitz and Yedid Hoshen 39377

Saying vs. Seeing: Evaluating LLM Conflict Awareness and Resolution Preference under Latent Claim-Evidence Disagreement
Daniel Hacker and Kfir Bar 39397

How Do Linear Probes Emerge? A Circuit-Tracing Framework with Concept-Targeted Attribution
Vedant Palit, Florent Draye, Terry Jingchen Zhang, Bernhard Schölkopf and Zhijing Jin . . . 39424

From Expressions to Needs: Hierarchical Interpretive Reasoning for Emotional Support Conversation
Xinjie Zhang, Haotian Fan, Zhaopei Huang, Shujie Ren, Wenxuan Wang and Qin Jin 39455

Risk-Adjusted Harm Scoring for Automated Red Teaming of LLMs in Financial Services
Fabrizio Dimino, Bhaskarjit Sarmah and Stefano Pasquali 39488

MIDR: Enrichment-Augmented Indexing for Multimodal Document Retrieval
Debanjan Mahata, Atharva Tendle, Daniel Preotiuc-Pietro, Yong Zhuang and Ozan Irsoy . . 39508

Unmasking Superficial Honesty: Mechanistic Head-Driven Activation Steering for Sycophancy-Free Medical VLMs
Mengliang Zhang, Prajwal Challa and Wentao Gao 39531

Can Persona-Conditioned Language Agents Simulate Counterfactual Online Discourse Dynamics?
Zhixiang Lu, Botao Zheng, Xueyuan Deng, Chong Zhang, Peiyu Hu, Mian Zhou and Jionglong Su 39545

Multi-Agent Evolve: LLM Self-Improve through Co-evolution
Yixing Chen, Yiding Wang, Siqi Zhu, Weixian Xu, Haofei Yu, Tao Feng, Muhan Zhang, Mostofa Patwary and Jiaxuan You 39563

Learning What to Retain: Gated-Memory Routing for Efficient Collaboration in Multi-Agent LLM Systems
Rakibul Hasan Rajib, Mengxin Zheng and Qian Lou 39594

MICE: Minimal Interaction Cross-Encoders for efficient Re-ranking
Mathias Vast, Victor Morand, Josiane Mothe and Benjamin Piwowarski 39612

Grounded Token Initialization for New Vocabulary in LMs for Generative Recommendation
Daiwei Chen, Zhoutong Fu, Chengming Jiang, Haichao Zhang, Ran Zhou, Tan Wang, Chunnan Yao, Guoyao Li, Rui Cai, Yihan Cao, Ruijie Jiang, Fedor Borisyuk, Jianqiang Shen, Jingwei Wu and Ramya Korlakai Vinayak 39629

Direct Diversity Optimization for Diverse Successful Trajectories in Preference Post-Training
Junwon Ko, Dong-Jae Lee, Minchan Kwon, Sunghyun Baek and Junmo Kim 39645

I Strongly Suspect This Website Is a Scam": Benchmarking PII Leakage and Detection without Defense in Autonomous Web Agents
Soham Roy, Sarthakbrata Halder, Arya Bharaty, Vaibhav Bhaskar, Yash Sinha, Dhruv Kumar, Srikant Panda and Murari Mandal 39664

Typological Feature Prediction with Large Language Models: An In-Context Learning Approach
Qianwen Wang, York Hay Ng, Aditya Khan and En-Shiun Annie Lee 39689

LPFQA: A Long-Tail Professional Forum-based Benchmark for LLM Evaluation
Liya Zhu, Peizhuang Cong, Jingzhe Ding, Aowei Ji, Wenya Wu, Jiani Hou, Chunjie Wu, Xiang Gao, Jingkai Liu, Zhou Huan, XueLei Sun, Yang Yang, Jianpeng Jiao, Liang HU, Xinjie Chen, Jiashuo Liu, Tong Yang, Zaiyuan Wang, Ge Zhang and Wenhao Huang 39705

Cross-Lingual LLM-Judge Transfer via Evaluation Decomposition
Ivaxi Sheth, Zeno Jonke, Amin Mantrach and Saab Mansour 39727

Rashid: A Cipher-Based Framework for Exploring In-Context Language Learning
Niyati Bafna, Ryan Soh-Eun Shim, Barbara Plank, David Yarowsky and Hale Sirin 39746

(How) Do MLLMs Report Bistable Images Like Humans?
Ryota Takatsuki, Tomoki Doi, Amane Watahiki, Anil K Seth and Hitomi Yanaka 39784

Stable Adaptive Thinking via Advantage Shaping and Length-Aware Gradient Regulation
Zihang Xu, Haozhi Xie, Ziqi Miao, Chen Qian and Lijun Li 39815

Tinted Frames: Question Framing Blinds Vision-Language Models
Wan-Cyuan Fan, Jiayun Luo, Declan Kutscher, Leonid Sigal and Ritwik Gupta 39834

Vision-Language Models Suppress Female Representations Under Ambiguous Input
Arnau Marin-Llobet, Simon Henniger and Mahzarin R. Banaji 39860

No One Model Catches Every Harm: Benchmarking Content Moderation Across Safety Scenarios
Afshin Oroojlooy and Hitesh Laxmichand Patel 39879

Subliminal Clocks: Latent Time Modelling in Diffusion Language Models
Maximo Rulli, Thomas V. Fontanari, Simone Petruzzi, Federico Alvetreti, Giorgio Strano, Donato Crisostomi, Giorgos Nikolaou, Tommaso Mencattini, Andrea Santilli, Emanuele Rodolà, Simone Scardapane and Alessio Devoto 39904

MoPLEx: Estimating Plackett-Luce Mixture Models for Multi-Objective Alignment
Dongyue Li, Ziniu Zhang, Lu Wang and Hongyang R. Zhang 39928

HSRM: Hidden-State Reward Models for Test-Time Verification
Xianzhi Li and Xiaodan Zhu 39947

From Token Probabilities to Semantic Constraints: Towards Declarative Probabilistic Evaluation of Language Models
Kyle Richardson, Cullen Anderson, Pranav Balakrishnan, Takuto Ban, Daksha Ladia, Ankita Gupta and Marisa Hudspeth 39963

LUMINA: LongitUdinal MultImodal reasoNing Agent for Medical Imaging
Zhaoqi An, Yunchang Xie, Xiaobing Yu, Yuyi Yang, Ahmad Chowdhury and Peijie Qiu . . 39983

The Masked Advantage: Uncovering Local-Language Access to Cultural Knowledge in LLMs
Yang Zhang, Xiao Fei, Amr Mohamed, Sarah Almeida Carneiro, Mersin Konomi, Mingmeng Geng, Ahmed Asaad, Guokan Shang and Michalis Vazirgiannis 40007

BenchBrowser: Retrieving Evidence for Evaluating Benchmark Validity
Harshita Diddee, Gregory Yauney, Swabha Swayamdipta and Daphne Ippolito 40024

ChildVox: A Speech, Audio, and Large Audio-Language Model Benchmark in Understanding and Characterizing Sound across Childhood
Tiantian Feng, Anfeng Xu, Xuan Shi, Aditya Kommineni, Shakhrul Iman Siam, Megan Micheletti, Zhonghao Shi, Helen Tager-Flusberg, Mi Zhang, Lynn K Perry, Catherine Lord, Daniel Messinger and Shrikanth Narayanan 40054

Same Payload, Different Channel: Measuring Trust Asymmetry in Tool-Using Language Models
Mohammed Sameer Syed and Rozhin Yasaei 40073

Bias Before Disclosure: A Behavioral Audit of LLM Assistants Under User-Identity Uncertainty
Hitesh Laxmichand Patel, Srikant Panda, Amit Agarwal, Tejaswini Kumar, Bhargava Kumar, Karan Dua, Jyotika Singh, Taki Hasan Rafi, Mithil Bangera, Yeshil Bangera and Gouthami Vadithya 40088

AVERT: Audio-Verified Adjudication for Spoken Dialogue State Tracking
Chunggi Lee and Hanspeter Pfister 40120

ParaRater: Enhancing Cross-Lingual Transfer in LLMs with Meta-Learning
Wenhan Han, Binbin Liu, Zhixun Chen, Yifan Zhang, Mykola Pechenizkiy, Meng Fang and Yin Zheng 40131

Using Prosody to Predict Syntactic Structure
Junghyun Min, Alex Warstadt, Tamar I Regev, Tiago Pimentel and Ethan Gotlieb Wilcox . 40145

GNN-CB: A Graph Neural Network Competition Benchmark for Human and LLM Evaluation
Murad Hossen, Tasneem Selim, Gurur Gamgam, Tuga Yousif, Abderrahmane Kasmi, Ikram Aissiou, Mubaraq Onipede, Faran Taimoor Butt, Sanae Zrigui, Rosa Y. G. Paccotacya-Yanque, Ignatius Balayo, Ikram Elhouiti, Hadil Affes, Bijay Adhikari, Sargam Goyal, Muhammad Ibrahim Isah, Mohammad Idrees Bhat, Samuel Kangoni Matia, Peguy Kem-Meka Tiotsop Kadzue, Maha Trabelsi, Emmanuel Owusu, Vinit, Nour Majdoub, Tamiru Alemnew and Islem Rekik 40160

Are We There Yet? Assessing Computer-Use Agents for Blind Users' Accessible Interaction with Desktop Applications
Satwik Ram Kodandaram, Monalika Padma Reddy, Xiaojun Bi, Jiawei Zhou, I. Ramakrishnan and Vikas Ashok 40189

PRISM: Agentic Retrieval with LLMs for Multi-Hop Question Answering
Md Mahadi Hasan Nahid and Davood Rafiei 40210

Seeing Through the MiRAGE: Evaluating Multimodal Retrieval Augmented Generation
Alexander Martin, William Gantt Walden, Reno Kriz, Dengjia Zhang, Kate Sanders, Eugene Yang, Chihsheng Jin and Benjamin Van Durme 40231

Litmus (Re)Agent: A Benchmark and Agentic System for Predictive Evaluation of Multilingual Models
Avni Mittal, Shanu Kumar, Sandipan Dandapat and Monojit Choudhury 40255

Span-Based Reasoning with Anonymous Knowledge Graphs
Daniel Guzman Olivares, Lara Quijano and Federico Liberatore 40283

Predicting Long-Term Conversational Skill Development in Mental-Health Crisis Counselors
Vivian Nguyen, Lillian Lee, Elizabeth A. Olson and Cristian Danescu-Niculescu-Mizil . . . 40304

EmoLASP: Emotion Recognition with Language Models and Answer Set Programming
Thao Le and Michael Thielscher 40322

VGI-Bench: Probing Visual Intelligence in Video Generation Models
Xuan He, Cong Wei, Yuhao Cheng, Linrui Ma, Yuxuan Zhang, Zuojun Li, Yuhao Wen, Jize Jiang, Zeyi Liu, Yuren Hao, Songcheng Cai, Keming Wu, Penghui Du, Kai Zou, Rui Yang, Chenkai Sun, Ke Yang, Ping Nie, Kelsey R Allen, Chenglong Wang, Michel Galley, Jianfeng Gao and ChengXiang Zhai 40335

The Anatomy of an ASR Hallucination
Hamees Sayed, Apoorv Singh, Kumar Aman and Akshat Mandloi 40372

FIDES: Faithful Inference via Deep Evidence Signals for Retrieval-Memory Conflict in RAG
Zhe Yu, Wenpeng Xing, Tiancheng Zhao, Mohan Li, Changting Lin and Meng Han 40386

HARP: Hadamard-Preconditioned Adaptive Rotation Processor for Extreme LLM Quantization
Artur Zagitov, Gleb Molodtsov and Aleksandr Beznosikov 40406

UNIQUE: Universal Top-k Sparse Attention for Training-free Inference and Sparsity-aware Training
Keqi Deng, Shaoshi Ling, Ruchao Fan and Jinyu Li 40428

QTEA: Ternary LLMs with Sparse Residual Salient Weight and By-Column Optimization
Yipin Guo, Jie Fu, Arun M George, Tareq Mahmoud and Siddharth Joshi 40441

Conformal Certification of Reasoning Trace Prefixes
Matt Y. Cheung, Ashok Veeraraghavan, Hanjie Chen and Guha Balakrishnan 40454

MABPD: Multi-Agent Bias Probing & Detection via Structured Argument Debate
Garvit Joshi, Stavya Dhyani, Jasmine and Arun Chauhan 40482

Breaking the Likelihood Trap: Variance-Calibrated Modulation for Large Language Model Decoding
Yuanhao Ding, Meimingwei Li, Esteban Garces Arias, Matthias Aßenmacher, Christian Heumann and Chongsheng Zhang 40502

DialToM: A Theory of Mind Benchmark for Forecasting State-Driven Dialogue Trajectories
Neemesh Yadav, Palakorn Achananuparp, Jing Jiang and Ee-Peng Lim 40529

The Consistency Illusion: How Multi-Agent Debate Hides Reasoning Misalignment
Xiaoyang Wang and Christopher Yang 40561

On the Role of Citations in Preference Data
Yu Hou, Hal Daumé III, Rachel Rudinger and William Gantt Walden 40582

Does VLA Even Know the Basics? Limitations in Commonsense and World Knowledge of VLA Models
Nikita Kachaev, Andrey Moskalenko, Matvey Skripkin, Nikita Kurlaev, Daria Pugacheva, Albina Burlova, Mikhail Kolosov, Denis Shepelev, Andrey Kuznetsov, Elena Tutubalina, Aleksandr Panov, Alexey Kovalev and Vlad Shakhuro 40600

Equity with Efficiency: An Empirical Study of Tokenizers for Multilingual Large Language Models
Kieron Seven Lee Jun Wei, Muhammad Reza Qorib, Andrew Ivan Soegeng and Hwee Tou Ng 40622

Do Value Vectors in Deep Layers Need Context from the Residual Stream?
Muyu He, Yuchen Liu, Qingya Huang and Li Zhang 40637

Dynamic Context Allocation for Concurrent Multi-Agent LLMs
Junho Na, Daiki E. Matsunaga, Nam Ho Koh, Jongmin Lee, Pascal Poupart and Kee-Eung Kim 40651

Model Editing for Targeted Translation Error Repair
Danna Zheng, Yang Liu, Zhu Liu, Vimal Bhat, Manoj Aggarwal, Gerard Medioni and Najmeh Sadoughi 40684

TokenDecouple: Frequency-aware Token Merging and Pyramidal Text Decay
Xin Qiu and Jialingan 40718

Beyond Referring Expressions: Scenario Comprehension Visual Grounding
Ruozhen He, Nisarg A Shah, Qihua Dong, Zilin Xiao, Jaywon Koo and Vicente Ordonez . 40742

AgentSocialBench: Evaluating Privacy Risks in Human-Centered Agentic Social Networks
Prince Zizhuang Wang and Shuli Jiang 40763

Where to Steer: Input-Dependent Layer Selection for Steering Improves LLM Alignment
Soham Gadgil, Chris Lin and Su-In Lee 40786

VLA Grounder: Language-Conditioning Space Optimization for Black-Box VLA Models
Damir Shodiev, Nikita Kachaev, Alexey Kovalev, Aleksandr Panov and Aleksei Staroverov 40818

When Structure Is Hidden: Measuring and Diagnosing the Compositional Gap in LLMs
Shravani Hariprasad, Mohsin Ali Mohammed, Vinija Jain and Aman Chadha 40834

Talk in Pieces, See in Whole: Disentangled and Hierarchical Representation Learning in Language-based Object Detection
Sojung An, Kwanyong Park, Yong Jae Lee and Donghyun Kim 40850

Are Single-Token Sparse Autoencoder Features Causally Necessary? Layer-Depth and SAE-Family Effects
Seonglae Cho, Zekun Wu, Kleyton Da Costa, Rishi Kalra, Ilham Wicaksono and Adriano Koshiyama 40875

SCAR: Shapley Credit Assignment for More Efficient RLHF
Meng Cao, Shuyuan Zhang, Xiao-Wen Chang and Doina Precup 40898

Drift No More: Spectral Locking for Long-Horizon Open-Weight Multi-Agent AI
Vishal Pramanik, Maisha Maliha, Susmit Jha and Sumit Kumar Jha 40916

Lost in Speech: Benchmarking, Evaluation, and Parsing of Spoken Bilingual Conversational Language Beyond Standard UD Assumptions
Nemika Tyagi, Olga Kellert, Holly Hendrix, Nelvin Licona-Guevara, Justin Mackie, Phanos Kareen, Megan Michelle Smith, Tatiana Gallego Hernandez, Samhitha Harish and Chitta Baral . . . 40948

PubMedCausal: A Span-Level Annotated Corpus for Causal Relation Extraction in Biomedical Text
Ifeoluwa Oluwatosin Kunle-John, Paul Josiah, Oluwatosin Agbaakin, Peter Olusegun Aina, Ikenna Odezuligbo and Sydney Anuyah 40966

Groundhog Bit-Flip Attack: Seeding Infinite Generation Loops in Mixture-of-Experts LLMs through Bit Flips
Huakang Lin, Tiancheng Zheng, Mingxuan Sun, Tianhong Xu, Yunsi Fei, Fan Terry Zhang and Ruyi Ding 40998

CLASP: Collaborative Latent Alignment from Semantic Preference Descriptions for Natural Language based Recommendation
Dilan Dinushka, Antoine Ledent and Hady W. Lauw 41013

Multi-Level Implicit Discourse Relation Classification via Sequence-Level Reinforcement Learning
Wei Liu, Xiaoyu Shen and Xiyan Fu 41027

Layer-Order Inversion: Rethinking Latent Multi-Hop Reasoning in Large Language Models
Xukai Liu, Ye Liu, Jipeng Zhang, Yanghai Zhang, Kai Zhang and Qi Liu 41041

Same Answer, Different Representations: Hidden instability in VLMs
Farooq Ahmad Wani, Alessandro Suglia, Rohit Saxena, Aryo Pradipta Gema, Wai-Chung Kwan, Fazl Barez, Maria Sofia Bucarelli, Fabrizio Silvestri and Pasquale Minervini 41059

TRACE: A Long-Term Conversational Recommendation Dataset with Evolving User Personas
Jihyung Ahn, Seonah Han, Chanho Park, Hyeongwoo IM, Daeyeong Kim, Subin Kim and Buru Chang 41088

Syntactic Belief Update as the Driver of Garden Path Processing Difficulty
Alan Zhou, Miloš Stanojević and John T. Hale 41137

AffectDF: The Most Comprehensive Benchmark for Speech Deepfake Detection against Emotionally Expressive Attacks
Aurosweta Mahapatra, Xiutian Zhao, Shreeram Suresh Chandra, Zihan Zhang, Zongyang Du, Ismail Rasim Ulgen, Kong Aik Lee, Nicholas Andrews, Carlos Busso and Berrak Sisman 41157

Narrative Function Patterns in LLM-Generated Chinese Web Fiction: A Proppian Analysis
Yuanchi Ma, Jiamou Liu, Kaize Shi, Hui He, Gang Zhang, Zijun Zhao, Renfen Hu and Maosong Sun 41179

Per-Step Conformal Calibration Does Not Transfer to Agent Trajectories
Li Ning, Yang Liu, Dongxin Guo, Yu Wu, SM Yiu and Jikun Wu 41197

Rent-a-RAG: Embedding-Space Watermarks for Auditing Third-Party RAG
Alexandr Goultiaev Tolstokorov, Kyriakos Mouratidis, Javad Dogani and Nikolaos Laoutaris 41216

Can Large Language Models Forecast What Researchers Study Next?
Fenghai Li, Zihan Tang, Haofei Yu, Yining Zhao and Jiaxuan You 41244

Every LLM-as-Judge Aggregator Must Sacrifice a Fairness Axiom
Yang Liu, Dongxin Guo, Li Ning, Yu Wu, SM Yiu and Jikun Wu 41275

Scalable Circuit Learning for Interpreting Large Language Models
Naiyu Yin, Dennis Wei, Tian Gao, Amit Dhurandhar, Karthikeyan Natesan Ramamurthy and Yue Yu 41296

The Provenance Horizon: Short Regulated Text Resists Reliable Watermarking Under Editing
Yanmin Zhao, Jikun Wu, Tengjiao Zheng, Yang Liu, Dongxin Guo and SM Yiu 41316

Aligning the Query Space: Greedy Information Projection for Language Model Data Selection
Victor Ye Dong, Kuan-Yun Lee, Jiamei Shuai, Shengfei Liu, Yi Liu and Jian Jiao 41334

How LLMs Build Fictional Worlds: Setting and Narrative Space in AI-Generated Creative Storytelling
Katrin Rohrbacher, Björn Nieth, Emmanuelle Salin, Bjoern Eskofier and Michaela Mahlberg 41362

Tracing Contextual Override: Mediated Causal Attribution in Language Models
Zongye Hu, Weiqing Luo, Haofeng Zhang and Ziyi Huang 41385

Escaping the KL Agreement Trap in On-Policy Distillation
Haoran Xin, Anhao Zhao, Ying Sun, Jin Li, Xiaoyu Shen and Hui Xiong 41418

From Pixels to Places: A Systematic Benchmark for Evaluating Image Geolocalization Ability in Large Language Models
Lingyao Li, Runlong Yu, Qikai Hu, Bowei Li, Min Deng, Yang Zhou and Xiaowei Jia 41432

Linguistic Context Recodes Visual Representations in Vision-Language Models
Brian Song, Michael A. Lepori and Ellie Pavlick 41458

Auditing RAG Necessity: A Conformal Test for Source-Independent Answers
Dongxin Guo, Yang Liu, Jikun Wu and SM Yiu 41480

What Drives the Zero-to-One Jump? Diagnosing and Steering Hard-Prompt Learning in RLVR
Rui Zhu, Yumou Liu, Zhenzhe Zheng and Fan Wu 41504

Zero-Shot Tabular Learning via Hypergraph Reasoning with LLMs
Geon Lee, Kyungho Kim, Dongwon Choi, Juyeon Kim, Shinhwan Kang and Kijung Shin . 41521

When Must Faithful Chain-of-Thought Fail? An Information-Theoretic Bound and Its Empirical Limits
Jikun Wu, Dongxin Guo and SM Yiu 41537

CauSight: Learning to Supersense for Visual Causal Discovery
Yize Zhang, Meiqi Chen, Sirui Chen, Bo Peng, Yanxi Zhang, Tianyu Li and Chaochao Lu 41556

Improving Argument Saliency Coverage in Small LLMs for Long Legal Opinion Summarization via Sequence-Level Distillation
Mohamed Elaraby, Ahmed Elhady and Diane Litman 41581

Diffusion Large Language Models for Visual Speech Recognition
Jeong Hun Yeo, Chae Won Kim, Hyeongseop Rha and Yong Man Ro 41596

InCoder-32B-Thinking: Industrial Code World Model for Thinking
Jian Yang, Wei Zhang, Jiajun Wu, Junhang Cheng, Tuney Zheng, Fanglin Xu, William Grant, Linzheng Chai, Yaxin Du, Haau-Sing Li, Yizhi LI, Yan Xing, Chuan Hao, Ran Tao, Ruihao Gong, Aishan Liu, Zhoujun Li, Mingjie Tang, Chenghua Lin, Siheng Chen, Xin Zhao, Xianglong Liu, Ming Zhou, Bryan Dai and Weifeng Lv 41611

When Quantization Preserves Accuracy but Not Evidence: Explanation-Aware Post-Training Quantization for Medical LLMs
Yeji Kim, Mi-Young Kim and Randy Goebel 41629

Decoding Strategies for Diffusion-Based ASR: A Systematic Evaluation of Confidence-Based Thresholding
Jeong Hun Yeo, Minsu Kim, Hyeongseop Rha and Yong Man Ro 41652

The Language of Evidence: Robustness and Bias in Code-Switched RAG
Medha Hira, Joel Mire and Maarten Sap 41664

LogicInception: Discourse-Level Indirect Prompt Injection via Causal Chain Construction
Mingshuai Liu, Yijie Peng, Chen Dongxian, Mingfeng Li and Xin Li 41686

Toward End-to-End Open-Weight LLM-RBMT for Endangered Languages
Jared Coleman and Bhaskar Krishnamachari 41703

Adaptive Latent Block Insertion via Online Reward Optimization
Yewon Cho, Hyeongsoon Hwang, Yein Park, Hyeon Hwang and Jaewoo Kang 41715

LiveSim: Simulating Environment-Shaped Users in Multi-Agent Live-Stream Ecosystems
Jiaqi Xu, Yiran Qiao, Jing Chen, Qiwei Zhong, Xiang Ao and Xueqi Cheng 41737

Analyzing LLM Reasoning to Uncover Mental Health Stigma
Sreehari Sankar, Aliakbar Nafar, Mona Barman, Hannah K. Heitz, Ashwin Kumar, Pouria Tohidi, Dailun Li, Danish Hussain, Russell DuBois, Hamed Hasheminia and Farshad Majzoubi 41757

Rethinking Diffusion Large Language Models via Time-aware Weighting and Block-oriented Refinements
Zheyong Xie, Shengrui Li, Li Zhong, Kaiyan Zhao, Fangcheng Shi, Jiakang Wang, Huayi Jin, Zhaokai Luo, Fei Zhao, Haifeng Liu, Jieying Ye, Chonggang Lu, Wenxuan Huang, Yue Liu, Zijie Meng and Shaosheng Cao 41778

LangMARL: Natural Language Multi-Agent Reinforcement Learning via Semantic Credit Assignment
Huaiyuan Yao, Longchao Da, Xiaoou Liu, Charles Fleming, Tianlong Chen and Hua Wei . 41793

One Form to Transfer Them All: Pretraining Multilingual Language Models Beyond Native Orthography
Muge Zhang, Aaron Jencks, Krishna Badikela, Yulia Tsvetkov and Sachin Kumar 41808

Confess What You Know: Forget-Set Misalignment with Model Knowledge in LLM Unlearning
Miso Kim, Georu Lee, Seungwon Jeong and Woojin Lee 41825

RoCo-ACE: Rollout-Conditioned Online Distillation for Retention-Aware Knowledge Injection
Yan Hong, Wei Li, Kedong Xiu, Jun Lan, Shuheng Zhou, Zhongcai Lyu, Huijia Zhu, Weiqiang Wang and Jianfu Zhang 41847

Learning Query-Specific Rubrics from Human Preferences for DeepResearch Report Generation
Changze Lv, Jie Zhou, Wentao Zhao, Jingwen Xu, Shihan Dou, Zisu Huang, Muzhao Tian, Xiaohua Wang, Yang Liu, Pluto Zhou, Tao Gui, Le Tian, Zhou Xiao, Xiaoqing Zheng, Xuanjing Huang and Jie Zhou 41870

Physiognomy Revisited: How Vision–Language Models Encode Sensitive Attributes Through Image Transformation
Soumyajit Datta, Arka Dutta and Ashiqur R. KhudaBukhsh 41898

Lessons Without Borders? Evaluating Cultural Alignment of LLMs Using Multilingual Story Moral Generation
Sophie Wu and Andrew Piper 41915

Benchmarking Language Models for Statistical Problem Formulation
Chen Wang, Junzhe Zhao, Xin Cong, Wanlu Deng and Ke Deng 41943

TDD-Agent: Test-Driven Reasoning for Code Generation
Hongyue Yu, Kefan Li, Jiakun Li, Hongzheng Chai, Yuan Yuan, Rui He and Junyi Wei . . . 41963

Contrastive Neuron Regulation for Chain-of-Thought Reasoning
Zitao Su, Xinyu Tang and Xin Zhao 41986

CAST: Critique-Aware Supervision for Training Reliable Long-Horizon Tool-Calling Agents
Amir Saeidi, Zehua Zhang, Rishitosh Kumar Singh, Naman Ahuja, Vivek Gupta, Ali Payani, Gaowen Liu, Jayanth Srinivasa and Chitta Baral 42006

SemPOI-RL: Aligning LLM Semantic Reasoning for Interpretable Out-of-Town POI Sequential Generation
Yunqi Liu, Yang Zhang, Ruixing Zhang, Liangzhe Han, Yi Qiao, Tongyu Zhu and Leilei Sun 42035

LLM Cache Bandits with Query Heterogeneity for Cost-Efficient Inference
Hantao Yang, Hong Xie, Xutong Liu and Defu Lian 42054

Race in the Record: Measuring Racial Signal and Model Inference in Police Arrest Narratives
Arka Dutta, Ted Forsyth, O. Nicholas Robertson and Ashiqur R. KhudaBukhsh 42088

DiffIE: Diffusion-based Open Information Extraction
Konstantin Fedorov and Valentin Malykh 42105

Seer: Efficient KV Cache Management for LLM Acceleration via Mitigating Semantic Search Inefficiency
Yiwei Hu, Fangxin Liu, Yue Liang, Xuan Wang, Cheng Li, Gezi Li, Li Jiang and Haibing Guan 42117

KPocBench: Benchmarking and Measuring Code Reasoning of Language Models in Linux Kernel
Ziyi Guo, Kexin Pei and Yan Chen 42131

PINGPONG: A Natural Benchmark for Multi-Turn Code-Switching Dialogues
Mohammad Rifqi Farhansyah, Hanif Muhammad Zhafran, Farid Adilazuarda, Maryam Ibrahim Mukhtar, Shamsuddeen Hassan Muhammad, Nedjma Ousidhoum, Sathwik Reddy, Ponnurangam Kumaraguru, Prashant Kodali, Sunayana Sitaram, Genta Indra Winata, Ayu Purwarianti and Alham Fikri Aji 42146

Can Your Agent Read the Source Faithfully? Benchmarking MLLM Agents on Source-Grounded Multimodal Retrieval
Wentao Gao, Shizhou Huang, Junling Zhuang, Advitya Garg, Mohammad Hasibur Rahman, Mengliang Zhang, Qin Qin and Yu Yu 42175

RealClawBench: Live OpenClaw Benchmarks from Real Developer-Agent Sessions
Zongwei Lv, Yaoming Li, Zhewen Tan, Yilun Yao, Yuxuan Tian, Lin Sun, Xiangzheng Zhang, Weihong Lin, Tong Yang and Guangxiang Zhao 42195

Beyond Naturalness: Probing Automated TTS Evaluators on Linguistically Grounded Dimensions
Oluwanifemi Bamgbose, Jash Shah, Simon Rosen, Lindsay Devon Brin, Hoang H Nguyen, Anke Koelzer, Rachel Hansen, Tara Bogavelli and Fanny Riols 42215

Inducing Task Models from Computer-Use Traces
Yucheng Jiang, Zora Zhiruo Wang, Ruishi Chen and Diyi Yang 42244

StreamUni: Achieving Streaming Speech Translation with a Unified Large Speech-Language Model
Shoutao Guo, Xiang Li, Mengge Liu, Chen Wei and Yang Feng 42267

From Inertia to Objectivity: Improving Deep Research Agents with Noise Isolation
Xiangxin Zhang, Zhanwei Zhang, Zhihang Fu, Binbin Lin and Wenxiao Wang 42282

GSM8K-V: Can Vision-Language Models Solve Grade School Math Word Problems in Visual Contexts?
Fan Yuan, Yuchen Yan, Yifan Jiang, Haoran Zhao, Tao Feng, Jinyan Chen, Yanwei Lou, Wenqi Zhang, Weiming Lu, Jun Xiao, Yueting Zhuang and Yongliang Shen 42302

Forewarned Is Not Forearmed: How Coordinated LLMs Shift Reader Opinions in News Comment Sections
Wooyoung Go, Yujin Potter, Hyoungshick Kim, Seungjin Ryu and Yongdae Kim 42361

A Comparative Analysis of LLM Memorization at Statistical and Internal Levels: Cross-Model Commonalities and Model-Specific Signatures
Bowen Chen, Namgi Han and Yusuke Miyao 42392

PACE: Towards Surfacing Hidden Conflicts in User Requests
Yoojin Kim, Jihyoung Jang and Hyounghun Kim 42421

SPARC-RAG: Adaptive Sequential–Parallel Scaling with Context Management for Retrieval-Augmented Generation
Yuxin Yang, Gangda Deng, Ömer Faruk Akgül, Nima Chitsazan, Yash Govilkar, Akasha Tigalappanavara, Shi-Xiong Zhang, Sambit Sahu and Viktor Prasanna 42480

Can LLMs Follow Medical Expert Logic? A Benchmark for Hierarchical Logical Consistency in Risk-of-Bias Assessment
Jiayu Huang, Zichen Tang, Qianhui Ling, Zemin Kuang and Haihong E 42509

Over-Refusal and Representation Subspaces: A Mechanistic Analysis of Task-Conditioned Refusal in Aligned LLMs
Utsav Maskey, Mark Dras and Usman Naseem 42550

CuRLRank: Adaptive Curriculum Reinforcement Learning for Reasoning-Intensive Reranking
Xin Kang, Yunxiao Zhao, Jiayang Zhang, Juncai Li, Zhichao Yan, XiaoQi Han, Ru Li and Delvin Ce Zhang 42563

PREREVIEWBENCH: Benchmarking LLMs for Editorial Desk Screening in Scholarly Publishing
Ming Chen, Zhiyuan Ning and Yi Du 42593

GLARE: Repository-Level Code Optimization Agents for Heterogeneous CPU-GPU Systems
Junfeng Gong, Cheng Liu, Junying Chen, Xiaowei Li and Huawei Li 42615

Faithful to the Frame: Source-Framing Propagation in LLM-Agent Decision Workflows
Yi-Ting Chiu and Chung-Chi Chen 42630

Can AI Agents Answer Your Data Questions? A Benchmark for Data Agents
Ruiying Ma, Shreya Shankar, Ruiqi Chen, Adit Chintamaneni, Yiming Lin, Sepanta Zeighami, Rajoshi Ghosh, Abhinav Gupta, Anushrut Gupta, Tanmai Gopal and Aditya Parameswaran 42651

CLEAR: Context Augmentation from Contrastive Learning of Experience via Agentic Reflection
Linbo Liu, Guande Wu, Han Ding, Yawei Wang, Qiang Zhou, Yuzhe Lu, Zhichao Xu, Huan Song, Panpan Xu and Lin Lee Cheong 42678

The Alignment Veto: How Safety Training Suppresses Cultural Knowledge in LLMs
Pardis Sadat Zahraei, Gokhan Tur, Dilek Hakkani-Tür and Ehsaneddin Asgari 42697

Beyond Single-Attribute Auditing: A Configurable Multi-Attribute Benchmark for Intersectional Fairness in LLMs
Zheng Zhang, Yifan Wu, Yuxiang Zhou, Qi Liu, Qingyang Mao, Ning Li, Weibo Gao, QiYuan Shi, Zhenya Huang and Baosheng Yu 42742

Detect, Remask, Repair: Diffusion Editing for Faithful Summarization of Evolving Contexts
Hao Zou, Zachary Horvitz, Chandhru Karthick, Zhou Yu and Kathleen McKeown 42763

From Instance Selection to Fixed-Pool Data Recipe Search for Supervised Fine-Tuning
Haodong Wu, Jiahao Zhang, Lijie Hu and Yongqi Zhang 42785

Features in Motion: Periodicity, Homeostasis, and Adaptivity in LLM Few-Shot Prompting
Ikhyun Cho, Sewoong Lee and Julia Hockenmaier 42819

ZipBench: Low-Cost Framework for Compressing Comprehensive Benchmarks of Large Language Models
Zhongzhan Huang, Junxin Li, Guoming Ling, Yupei Lin, Shanshan Zhong and Hefeng Wu 42839

JudgeSQL: Reasoning over SQL Candidates with Weighted Consensus Tournament
Jiayuan Bai, Xuanguang Pan, Chongyang Tao and Shuai Ma 42870

PaSBench-Video: A Streaming Video Benchmark for Proactive Safety Warning
Yusong Zhao, Yuejin Xie, Youliang Yuan, Junjie Hu, Juning Guo, Yujiu Yang and Pinjia He 42889

Group Adaptive Clipping Policy Optimization
Sheng Jia, Xiao Wang, Shiva Kasiviswanathan and Rein Houthooft 42914

Lexical Completion from Visual Evidence for Multimodal Topic Modeling
Dayu Guo, Zhiwen Luo, Nizar Bouguila and Wentao Fan 42929

Kurtosis-Aware Coupled Sparsity: Training-Free Activation Sparsity for Large Language Models
Guangyan Li, Luoyang Sun, Siyuan Li, Haifeng Zhang, Yongqiang Tang and Wensheng Zhang 42942

Policy and World Modeling Co-Training for Language Agents
Ning Lu, Baijiong Lin, Shengcai Liu, Jiahao Wu, Haoze Lv, Yanbin Wei, Lingting Zhu, Shengju Qian, Xin Wang, Ying-Cong Chen, Qi Wang and Ke Tang 42962

What Makes Agent Memory Useful for Reliable Unanswerable Question Handling?
Chuanyuan Tan, Junjie Yu, Yuxin Wang, Yining Zheng, Xipeng Qiu and Wenliang Chen . . 42980

DCFA: Dual-view Causal-inspired Attribution for Failure Reasoning in LLM-based Multi-agent Systems
Zehao Wang, Lanjun Wang, Shilong Jin, Junjie Chen and Yanghua Xiao 42999

Towards Reliable Long-Context Reasoning: Detecting Unanswerable Questions via FactGuard
Qian-Wen Zhang, Biao Liu, Fang Li, Jie Wang, Lingfeng Qiao, Yifei Yu, di Yin and Xing Sun 43026

BrowseComp-Verify: Benchmarking Self-derived Verification Agents in Web Search
Guanzhen Li and Shijie Li 43046

A Better Spur Should Start From Each Objective
Shanwen Mao, Hao Zhang, Guangtao Nie, Zhiheng Li, Huimu Wang, Sulong Xu and Gu Simiu 43076

PGMem: Tightly Coupled Persona–Memory Graph for Lifelong Personalized Agents
Wonjun Choi, Yerim Kim, Yukyung Lee and Susik Yoon 43087

DMV-Bench: Diagnosing Long-Horizon Multimodal Agents' Visual Memory with Incidental Cue Injection
Yujin Tang, Chenming Shang, Ruize Xu and Nikhil Singh 43113

HarnessForge: Joint Harness and Policy Evolution for Adaptive Agent Systems
Mingju Chen, Can Lv, Guibin Zhang, Heng Chang and Shiji Zhou 43130

BLUE: Toward Better Language Use in Efficient Vision-Language-Action Models for Autonomous Driving
George Ling, Lijin Yang, Hao Yang and Zhongzhan Huang 43156

The Flaw of Averages: Measuring Benchmark-Level Distributional Robustness
Arda Uzunoglu, Tianjian Li and Daniel Khashabi 43186

AfriSUD: A Dependency Treebank Collection for Evaluating Models on African Languages
Happy Buzaaba, Cheikh Mouhamadou Bamba Dione, David Ifeoluwa Adelani, Sylvain Kahane, Bruno Guillaume, Kim Gerdes, Shamsuddeen Hassan Muhammad, Kevin Guan, Anuoluwapo Aremu, Naome A Etori, Utitofon Inyang, Peter Nabende, David Sabiiti Bamutura, Andiswa Bukula, Idris Akinade, Chinedu Uchechukwu, Rooweither Mabuya and Christiane Fellbaum 43227

Beyond Static Retrieval: Opportunities and Pitfalls of Iterative Retrieval in GraphRAG
Kai Guo, Xinnan Dai, Shenglai Zeng, Harry Shomer, Haoyu Han, Yu Wang and Hui Liu . . 43249

STILL: Recovering Lowered STL Semantics for LLM-assisted C++ Decompilation
Xiaohan Wang and Kevin Leach 43272

NARRATOR: Narrative-Aware Routing and Evaluation of Large Language Models
Ahmed Hamdi, Jose G Moreno, Antoine Doucet and Fabio A. González 43288

LTEM: LLM-Friendly Table-Enhanced Markdown for Table Understanding and Generation
Eunbin Kim, Hyeonseok Kang and Sangkeun Jung 43304

HyperGVL: Benchmarking and Improving Large Vision-Language Models in Hypergraph Understanding and Reasoning
Yanbin Wei, Chun Kang, Siwei Li, Haoxuan Che, Yang Chen, Hua Liu, Jian Liu, Zhuang Liu, Can Ouyang, Fei Xing, Lei Sha, Rui Liu, Yu Zhang and James Kwok 43325

Countdown-Code: A Testbed for Studying The Emergence and Generalization of Reward Hacking in RLVR
Muhammad Khalifa, Zohaib Khan, Omer Tafveez, Hao Peng and Lu Wang 43345

TIDE: Tree-Induced Fine-Grained Diagnostic Evaluation from Model Behaviors
Jiayi Chen, Hui Li, Yujiong Shen, Qiyuan Peng, Yanxin Li, Chenhao Huang, Shaofan Liu, Guoqiang Zhang, Junjie Ye, Shihan Dou, Ming Zhang, Maxm Pan, Tao Gui, Qi Zhang and Xuanjing Huang 43364

Learning Interaction-Centric Capability Representation for LLM Tool Calling
Hanyi Wang, Ruihan Jin, Jinyang Wu, Zhengqi Wen and Jianhua Tao 43394

CCStories: Improving the Safety and Age-Appropriateness of LLM-Generated Chinese Children's Stories
Jiaxu Zhao, Qin Zhang, Xiaoxing Wang, Liang Ding, Yulong Pei, Xiaojun Chen, Meng Fang, Tanja Käser and Mykola Pechenizkiy 43410

Lost at the End: Primacy Bias in Multimodal Retrieval-Augmented Question Answering
Jieyuan Liu, Jianyang Gu, Shijie Chen, Jefferson Chen and Zhen Wang 43430

Debate-aware Learning Makes LLMs Genuine Debaters
Yufei Zhang, Kai Xiong, Yuxiong Yan, Bibo Cai, Xiao Ding, Bing Qin and Ting Liu 43450

Policy of Thoughts: Scaling LLM Reasoning via Test-time Policy Evolution
Zhengbo Jiao, Hongyu Xian, Qinglong Wang, Yunpu Ma, Zhebo Wang, Dezhang Kong and Meng Han 43472

EQUAL: Evaluating Pluralistic Alignment of LLMs Across Educational Perspectives
MD Azizul Hoque, Mark Dras and Usman Naseem 43491

Efficient Token-Level Policy Optimization for Masked Diffusion Models via Velocity Alignment
Chenxing Wei, Jiazhen Kang, Hong Wang, Jianqing Zhang, Hao Jiang, Xuxiaolong, Ningyuan Sun, Ying He, Fei Yu, Yao Shu and Bo Jiang 43516

Help or Hurdle? Rethinking Model Context Protocol-Augmented Large Language Models
Wei Song, Haonan Zhong, Ziqi Ding, Zhenchang Xing, Yulei Sui, Jingling Xue and Yuekang Li 43537

IdioLink: Retrieving Meaning Beyond Words Across Idiomatic and Literal Expressions
Kai Golan Hashiloni, Daniel Fadlon, Lior Livyatan, Ofri Hefetz Shahar, Jiahuan Pei and Kfir Bar 43554

Cultural Moment Benchmark: Evaluating Video Cultural Reasoning and Grounding in Southeast Asia
Burak Satar, Zhixin Ma, Cheng Yu-Tong, Huy Hoang Tran, Phuong Anh Nguyen and Chong-Wah Ngo 43574

Beyond Codebook Uniformity: Triplet-Level Metrics and Triplet-Driven Semantic IDs for Generative Retrieval
Xiang Tan, Dongliang Liao, Junwu Du, Haijun Wu, Run He, Shuquan Man, Ziqian Zeng and Huiping Zhuang 43608

How Robust Are Automated Fact-Checking Systems? A Cross-Benchmark Evaluation
Aida Usmanova, Zangir Iklassov, Markus Leippold and Ricardo Usbeck 43623

Can Workflow Search Be Amortized? Demonstration-Guided Synthesis of LLM Agent Workflows
Shiyi Du, Jiayuan Liu, Weihua Du, Yue Huang, Jiayi Li, Yingtao Luo, Xiangliang Zhang, Vincent Conitzer and Carl Kingsford 43645

X^3-OPD: Distilling Reasoning into Large Audio-Language Models via On-Policy Alignment
Dongjie Fu, Di Cao, Zihan Zhang, Wenxu Jia, Yifu Chen, Shengpeng Ji, Yu Zhang, Xize Cheng and Tao Jin 43674

Beyond Case-Level Similarity: Multi-Defendant Legal Case Retrieval via Individual-Aware Reasoning
Yi Yang, Ruize Ma, Yi Feng, Chuanyi Li and Vincent Ng 43688

REFINE: LLM Refinement over Budgeted Text-Attributed Graphs for Personalized Medical Concept Representation
Mohsen Nayebi Kerdabadi, Arya Hadizadeh Moghaddam, Dongjie Wang and Zijun Yao . . 43714

Beyond Append-Only Decoding: A Recursive Edit-Based Encoder for Arithmetic Length Generalization
Bong Gyun Kang, Junyong Ahn, Jieun Byeon and Sungroh Yoon 43731

How Do Document Parsers Break? Auditing Structural Vulnerability in Document Intelligence
Yue Chen, Yihao Wang, Ziyi Tang, Yongsen Zheng and Keze Wang 43751

RePair: Turning Retrieval Failures into Counterfactual Hard Pairs
Siyi Liu, Xiaorong Zhu, Enjun Du, Xinyu Zuo, Lisheng Duan, Haijin Liang, Jin Ma, Junfu Pu and Yongqi Zhang 43774

TimeBlind: A Spatio-Temporal Compositionality Benchmark for Video LLMs
Baiqi Li, Kangyi Zhao, Ce Zhang, Chancharik Mitra, Jean De Dieu Nyandwi and Gedas Bertasius 43801

Rethinking Clustering as a Conversational Supervision Problem
Theo Rode, Aimee Co, Qingyang Zeng, Sanaya Gandhi, Ananya Purwar, Jonathan P. Chang, Masoud Hashemi, Khyati Mahajan, Filip Deryckere and Fabio Casati 43823

TouchThinker: Scaling Tactile Commonsense Reasoning to the Open World with Large-scale Data and Action-aware Representation
Kailin Lyu, Di Wu, Pengwei Zhang, Yuhang Zheng, Yingxin Lai, Long Xiao, Kangyi Wu, Pengna Li, Chen Gao, Lianyu Hu, Xiaobin Hu, Jie Hao, Ce Hao, Weihao Yuan and Shuicheng Yan 43841

SkillAttack: Automated Red Teaming of Agent Skills through Attack Path Refinement
Zenghao Duan, Yuxin Tian, Zhiyi Yin, Liang Pang, Jingcheng Deng, Wenbin Duan, Yuyao Ge, Zihao Wei, Shicheng Xu and Xueqi Cheng 43859

Skim and Skip: Hierarchical Adaptive Inference for Efficient Multimodal Retrieval
Meng Gao, Yizhen Zhang, Yang Ding, Ziqi Dai, Shuoshuo Zhang, Junjie Wang, Taiqiang Wu, Chufan Shi, Lei Ji, Jian Jiao, Linfeng Zhang, Yeyun Gong and Yujiu Yang 43873

GR^2: Mitigating Granularity Collapse via Adaptive Query Routing in Generative Retrieval
Soona Hong, Dohyeon Lee, Jaeyoung Kim and Seung-Won Hwang 43891

CoSineVerifier: Agentic Verification for Computation-Oriented Scientific Questions
Ruixiang Feng, Zhenwei An, Yuntao Wen, Ran Le, Yiming Jia, Chen Yang, Zongchao Chen, Ke Shi, Lisi Chen, Shen Gao, Shuo Shang, Yang Song and Tao Zhang 43903

Calibrating Student Models via Uncertainty-Integrated Distillation
Ruixuan Zhang, Hongye Tan, Zhiwei Hu , XiaoQi Han, Ru Li, Jeff Z. Pan and Victor Gutierrez Basulto 43923

Call Neighbours Yourself: Graph Walks with Destination-Conditioned On-Policy Self-Distillation
Yilun Liu, Boyu Luo, Yanran Tang, Ruihong Qiu and Zi Huang 43939

Learning to Explore the Crux in Formal Theorem Proving
Linhan Li, Wenyao Cui, Gong-Duo Zhang, Lihong Gu, Lin Zhou, Jianyun Shang and Huaping Zhang 43958

Mitigating False Credit Propagation: Probabilistic Graphical Reward Aggregation for Rubric-Based Reinforcement Learning
Can Lv, Mingju Chen, Heng Chang and Shiji Zhou 43980

SEA-SpeechBench: A Large-Scale Multitask Benchmark for Speech Understanding Across Southeast Asia
Jingyi Liao, Wenyu Zhang, Zhuohan Liu, Yingxu He, Geyu Lin, Xunlong Zou, Shuo Sun, Syed Ali Redha Alsagoff and AiTi Aw 44004

VIGOR: Outcome-Contrastive Visual Grounding for Multimodal Policy Optimization
Haohao Xu, Yuyao Ge, Qiuyue Zhang, Zhaoli Deng, HongLin Liu, Sheng Ang, Likang Wu, Shenghua Liu and Hongke Zhao 44032

From Saliency to Discriminability: Rank-Preserving Visual Token Pruning for VLM Rerankers
Siyi Liu, Hanjun Yang, Chenchen Zhang, Xiaorong Zhu, Xinyu Zuo, Lisheng Duan, Haijin Liang, Jin Ma, Junfu Pu and Yongqi Zhang 44052

When Deep Research Agents Stagnate: Enhancing Reasoning with Retrieval-Aware Agent Control
Heydar Soudani, Elizabeth Lingg, Faegheh Hasibi and Navid Rekabsaz 44072

SpIDER: Spatially Informed Dense Embedding Retrieval for Software Issue Localization
Shravan S Chaudhari, Rahul Thomas Jacob, Jiajun Cao, Muhammad Shihab Rashid, Mononito Goswami and Christian Bock 44096

FigEx2: Visual-Conditioned Panel Detection and Captioning for Scientific Compound Figures
Jifeng Song, Arun Das, Pan Wang, Hui Ji, Kun Zhao and Yufei Huang 44132

RAD-PROTEIN: Multi-source Knowledge-enriched Retrieval-augmented Protein Structure Searching
Kexin Ding, Alison Biester, Brinda Vallat, Chenghua Shao, Dennis W Piehl, Dongwook Won, Joan Segura, Jose M. Duarte, Mu Zhou, Zukang Feng, Yana Rose, Stephen K. Burley and Dimitris N. Metaxas 44147

DataShield: Uncovering Risky Fine-Tuning Data Across LLMs Through Consensus Subspace Alignment
Zefeng Wu, Weiwei Qi, Jielong Chen, Tianhang Zheng, Di Hong, Chaochao Lu, Liang He, Zhan Qin and Kui Ren 44165

Evaluating the Evaluators: A Systematic Analysis of Reference Bias in Audio Captioning Metrics
Jaeheon Sim and Kyogu Lee 44191

LLMs Infer Cultural Context but Fail to Apply It When Responding
Yisong Miao, Jian Zhu and Vered Shwartz 44208

When Accuracy Aligns and Fails: Diagnosing Human–LLM Response Alignment at Population Scale
Seoyoon Park, Hyeji Choi, Subin An, Xiaonan Wang, Heejae Kim, Joeun Kang and Hansaem Kim 44234

Task–Specificity Score: Measuring How Much Instructions Really Matter for Supervision
Pritam Kadasi, Abhishek Upperwal and Mayank Singh 44255

Hybrid Linear Attention Done Right: Efficient Distillation and Effective Architectures for Extremely Long Contexts
Yingfa Chen, Zhen Leng Thai, Zihan Zhou, Zhu Zhang, Xingyu Shen, Shuo Wang, Chaojun Xiao, Xu Han and Zhiyuan Liu 44268

SIP: Aligning LLM-based TableQA with Table Semantics via SQL Logic
Zhen Yang, Ziwei Du, Minghan Zhang, Wei Du, Haodong Zou, Shicheng Tan, Jie Chen and Shu Zhao 44287

MiRD: Reliable Set-Valued Prediction for Open-Ended Question Answering via Miscoverage Risk Decomposition
Anqi Hu, Zijun Jia, Zhiyuan Wang and Bo Fu 44300

MSTree: Multi-Scale Evidence Tree for Training-Free Long-Form Video Question Answering
Chen Dongxian, Mingshuai Liu, Yijie Peng, Hengrui Liu and Xin Li 44321

Breaking Failure Cascades: Step-Aware Reinforcement Learning for Medical Multimodal Reasoning
Junha Jung, Minbyul Jeong, Suhyeon Lim, Sungwook Jung, Jaehoon Yun, Taeyun Roh, Mujeen Sung and Jaewoo Kang 44337

HiVe: Beyond Static Prompts for Multitask Learning via Hierarchy-based Vertical Mixture-of-Experts
Hyeonjik Bae, Minyeol Kim and Susik Yoon 44376

SkillShift-Bench: Benchmarking Skill Learning under Environment Evolution for Web Agents
Bo-Yu Chen, Zhi Zhou and Yu-Feng Li 44392

Semantic Flow Regularization: Teaching LLMs to Generate Diverse Yet Coherent Responses
Kerui Peng, Li Feifei, Xingyu Fan and Que Wenhui 44462

SkillForge: Learning from Success–Failure Boundaries for Consistent Long-Horizon Agents
Guangya Wan, Hao Wang, Xiao Ling, Mingyang Ling, Xiaoqi Ren and Sheng Li 44486

Unified-MAS: Universally Generating Domain-Specific Nodes for Empowering Automatic Multi-Agent Systems
Hehai Lin, Yu Yan, Zixuan Wang, Bo Xu, Sudong Wang, Weiquan Huang, Ruochen Zhao, Minzhi Li and Chengwei Qin 44511

Distance Is Not Enough: Forget-Retain Alignment Gap Predicts LLM Relearning Robustness
Yi Chen, Hanna Hsieh, Shuhong Liu, Chuanbo Hua, Zihan Ma, Kun Wang and Joo-Young Kim 44542

No Training, No Architecture Change: Per-Cell Warm-Up Calibration for LLM Agent Memory
Wenwen Liao and Yuansong Wang 44556

IESR: Efficient MCTS-Based Modular Reasoning for Text-to-SQL with Large Language Models
Tao Liu, Jiafan Lu, Bohan Yu, Pengcheng Wu, Liu Haixin, Guoyu Xu, Xiangheng Li, Lixiaoli, Jiaming Hou, Zhaoshijun, Xinglin Lyu, Kunli Zhang, Yuxiang Jia and Hongying Zan 44576

DITING: Evaluating Sentence-Level Cultural Meaning in Web Novel Translation
Enze Zhang, Jiaying Wang, Mengxi Xiao, Jifei Liu, Zhiwei Hu , Ziyan Kuang, Rui Dong, Youzhong Dong, Mohsinul Kabir, Victor Gutierrez Basulto, Min Peng, Qianqian Xie and Sophia Ananiadou 44605

Robust Multi-Agent LLMs under Byzantine Faults
Haejoon Lee, Vincent-Daniel Yun, Dimitra Panagou and Sai Praneeth Karimireddy 44639

Goldsmith: Gold-Loss-Guided Definition Optimization with an Agentic Annotation Harness
Yihan Li, Hanyi Zhang, Xiaoxi Jiang and Man Guo 44658

StructSynth: Dependency Graphs as Generation Plans for Low-Data Tabular Synthesis with Language Models
Siyi Liu, Yujia Zheng, Haoyang LI and Yongqi Zhang 44678

Rethinking Evaluation Practices in Automated Essay Scoring: A Quantitative Analysis over the Past Decade
Salam Albatarni, May Bashendy, Sohaila Eltanbouly, Marwan Sayed and Tamer Elsayed . . 44705

RAG over Thinking Traces Can Improve Reasoning Tasks
Negar Arabzadeh, Wenjie Ma, Sewon Min and Matei Zaharia 44722

Beyond Perfect APIs: Evaluating LLM Agents Under Realistic API Complexity
Doyoung Kim, Zhiwei Ren, Jie Hao, Zhongkai Sun, Lichao Wang, Xiyao Ma, Xu Han, Jun Yin, Heng Ji, Wei Shen, Xing Fan, Benjamin Z. Yao and Chenlei Guo 44744

ProcessLight: Process Supervision for Large Language Model Based Traffic Signal Control
Shuaitao Zhao, Tianlong Zhou, Weijie Wang, Shi Jiasheng and Weixiong Rao 44775

Process Rewarding via Solution-Space Alignment
ChenZhuo Zhao, Yue Huang, Pu Zhao, Fangkai Yang, Lu Wang, Lin Ma, Zhi Yang, Youling Huang, Xinda Wang, Junting Lu, Qingwei Lin, Saravan Rajmohan and Dongmei Zhang 44793

SampKV: Saliency-Aware Mixed-Precision KV Cache for Efficient Diffusion Language Model Inference
Zhenrui Wang, Haodong Chang, Yang Sui, Rongjian Liang, Bo Yuan and Jiang Hu 44808

MUA: Multiple Context Units Attention For Large Language Models
Zhiyi Li, Yaming Yang, Kai Shuang and Yiren Chen 44822

HAL: Inducing Human-likeness in LLMs with Alignment
Masum Hasan, Junjie Zhao and Ehsan Hoque 44834

Multi-Block Diffusion Language Models
Yijie Jin, Jiajun Xu, Yuxuan Liu, Chenkai Xu, Jiajun Li, Kai Yu, Pengfei Liu and Zhijie Deng 44855

Advancing DialNav through Automatic Embodied Dialog Augmentation
Leekyeung Han, Sangwon Jung, Hyunji Min, Jinseong Jeong, Minyoung Kim and Paul Hongsuck Seo 44876

The Cold-Start Safety Gap in LLM Agents
Chung-En Sun, Linbo Liu and Tsui-Wei Weng 44905

BridgeGap: Belief-Grounded Training for Medical Dialogue Agents
Weijun Li, Tao Yang, Jiayu Yang and Fei Chao 44932

The Past Is Not Past: Memory-Enhanced Dynamic Reward Shaping
Yang Liu, Enxi Wang, Yufei Gao, Weixin Zhang, Bo Wang, Zhiyuan Zeng, Yikai Zhang, Yining Zheng and Xipeng Qiu 44948

When Does Reasoning Age? Survival Analysis of Step-Level Error Hazard in LLM Chains
Raj Firke and Rajeswari Kannan 44972

RAMP: Adapting One Agent to Multiple Retrievers via Behavioral Probing
Weijun Li, Tao Yang, Jiayu Yang and Fei Chao 44989

Aligned but Not Partner-Specific: How Multimodal LLM Agents Succeed in Reference Games Without Forming Conceptual Pacts
Po-Ya Angela Wang, Chinmaya Mishra, Asli Ozyurek, Paula Rubio-Fernandez and Esam Ghaleb 45006

Locality-Aware Redundancy Pruning for LLM Depth Compression
Vincent-Daniel Yun, Youngrae Kim, Woosang Lim, YoungJin Heo, Minkyu Kim and Sunwoo Lee 45024

Why does Weak-OOD Help? A Further Step Towards Understanding Jailbreaking VLMs
Yuxuan Zhou, Yuzhao Peng, Yang Bai, Kuofeng Gao, Yihao Zhang, Yechao Zhang, Xun Chen, Tao Yu, Tao Dai and Shu-Tao Xia 45038

Bargaining with Internal Evidence: Receiver-Constrained Decoding for Factual Reliability in Large Language Models
Chenhao Lin, Mouyi Zhang, Peng Chen, Xin Zheng, Yuejia Hong and Nan Gao 45068

RoleKV - Role-Aware KV Cache Management for the Inverted Age-Importance of LLM Agent Context
Eashan Dash, Arun Ramachandran and Prakash Raghavendra 45082

First-Stage Groundwork Shapes Downstream RL Outcomes: Two-Stage Post-Training for LLM Personality Acquisition
Taisei Ozaki, Tetsuya Hashimoto, Miki Ota, Keisuke Hori, Zhenyu He and Mitsuhiro Mabuchi 45104

Perceive to Hypothesize, Verify to Ground: an Agentic Reasoning Framework for Open-World Geo-Localization
Yutian Jiang, Ruijie Li, Sisuo Lyu, Xixuan Hao, Qingxiang Liu, Yongzi Yu and Yuxuan Liang 45132

GLOBE: Training-Free Calibration for Image-Text Retrieval with Global-Local Bias and Graph Evidence
Xuzhe Zhao, Zhirun Yue, Xingyue Xu, Houde Liu and Kaichen Dong 45154

Learning to Retrieve: Dual-Level Long-Term Memory for Text-to-SQL Agents
Yibo Wang, Nikki Lijing Kuang, Philip S. Yu, Zhewei Yao and Yuxiong He 45174

Can Vision-Language Models Solve the Shell Game?
Tiedong Liu and Wee Sun Lee 45198

Benchmarking Correctness and Security in Multi-Turn Code Generation
Ruchit Rawal, Jeffrey Yang Fan Chiang, Chihao Shen, Jeffery Siyuan Tian, Aastha Mahajan, Tom Goldstein and Yizheng Chen 45222

GraphVerse: A Comprehensive Visual Graph Reasoning Benchmark for Multimodal Large Language Models
Yuanfu Sun, Yuanhang Ren, Kang Li, Chuanhao Ji, Jiaxi Li, Jiajin Liu, Ninghao Liu and Qiaoyu Tan 45259

Only Pay What You Must Spend: On-Demand Privacy Budget Payment for Differentially Private RAG
Zhonghao Sun, Zhiliang Tian, Xinyue Fang, Shuo Ma, Juhua Zhang, Yiping Song and Dongsheng Li 45293

The Illusion of Control: Why Bare Classifier Inversion Silently Fails in Concept-Bottleneck Text Generation
Qi Bing and Xiaowei Shao 45311

CARE: Contrastive Anchor-based Rubric Evolution for Large Language Model Post-Training
Siyuan Li, Xinxin Song, Chen Ruinian, Jingjing Fan, Tingxiong Xiao, Yangen Hu, Ke Zeng and Jinli Suo 45334

Human Psychometric Questionnaires Mischaracterize LLM Behavior
Woojung Song, Dongmin Choi, Yoonah Park, Jongwook Han, Eun-Ju Lee and Yohan Jo . . 45356

Learning to Think Like a Cartoon Captionist: Incongruity–Resolution Supervision for Multimodal Humor Understanding
Hatice Merve Vural, Doğa Kukul, Ege Erdem Özlü, Demir Ekin Arıkan, Bob Mankoff, Erkut Erdem and Aykut Erdem 45402

LatentRevise: Learning from Zero-Hit Reasoning
YiQiu Guo, Xueting Han, Qi Jia, Guangtao Zhai and Jing Bai 45443

After the Attacker Leaves: Benchmarking Carryover Attacks on LLM Agents
Zhiyue Xu, Xukai Wang, Zhiyuan Huang, Mingjie Zhan, Jian Zhao and Xu-Yao Zhang . . . 45456

MCTQ: Modality-Coordinated Transformation Quantization for Multimodal Large Language Models
Yingen Liu, Lanfeng Wu, Fan Wu, Wu Liyan, Ruihui Li, Zhuo Tang, Eng Siong Chng and Kenli Li 45477

ArcANE: Do Role-Playing Language Agents Stay in Character at the Right Time?
Woojung Song, Nalim Kim, Sangjun Song, Chaewon Heo, Jongwon Lim and Yohan Jo . . . 45495

FEA-SLT: A Gloss-Free End-to-End Framework for Facial-Expression-Aware Sign Language Translation
Guobin Tu and Di Weng 45546

LoRA-RG: Residual Gradient-Guided Rank Initialization and Expansion
Dayoung Kwon, Jaeseong Lee and Seung-Won Hwang 45571

CASG: Drive Text to IDs for Sequential Recommendation via Collaboration-Aware Spectral Guidance
Yichen Liu, Lianyue Zhang, Cong Xu, Zhangchi Zhu, Yanyu Chen, Yunhang He, Jun Wang and Wei Zhang 45588

Humans Introduce, Models Elaborate: Asymmetric Narrative Agency in Human-LLM Co-Writing
Halfdan Nordahl Fundal, Yuri Bizzoni, Charlotte Gjørup Bilde, Ida Bække Johannesen and Rebekah Brita Baglini 45603

Contrastive Attribution in the Wild: An Interpretability Analysis of LLM Failures on Realistic Benchmarks
Rongyuan Tan, Jue Zhang, Zhuozhao Li, Qingwei Lin, Saravan Rajmohan and Dongmei Zhang 45618

TYLER: Typed Latent Reasoning for Language Models —— When to Think, What to Compute, and How Much to Allocate
Hanyu Lin, Min Cai, Jiawei Wen and Haodi Zhang 45650

Where Hallucinations Live: A Cross-Architecture Circuit in VQ-Tokenized Vision-Language Models
Shamanthak Hegde, Xiangrui Liu, Maitreya Patel and YZ Yang 45671

MeRP: Efficient Multi-Expert Routing Prompt Encoding for Multilingual Translation with LLMs
Fengzhao Fan, Zhenhan Wang, Shengxiang Gao, Cunli Mao and Zhengtao Yu 45690

CLAAS: Conditional LVLM-enabled Attribute-Aware response Simulation for Public Health Education
Anh N Nhu, Jenna Yang, Laura Yu Zheng, Sylvette La Touche-Howard, Tracy M. Zeeger and Ming Lin 45706

Bridging Modality Disconnect in Self-Reflection via Closed-Loop Visually Grounded Verification
Haoyu Zhang, Zhi Gao, Pengxiang Li, Xintong Zhang, Yuwei Wu, Rui Gao, Mingyang Gao, Xuanyan Chen, Anqian Cen, Che Sun and Yunde Jia 45747

RVSD: Retrieval Vision Sparse Decoding for Mitigating Visual Hallucinations in Large Vision-Language Models
Canjie Liu, Jiawen Kang, Jinbo Wen and Zishao Zhong 45789

Behavior-Preserving KV Cache Compression
Doo Hwan Hwang, Junyoung Jang, Junho Na, Hosung Lim and Kee-Eung Kim 45807

By Their Fruits You Will Know Them: Comparing Formalizations of Law by the Decisions They Encode
Julius Vernie and Matthias Grabmair 45830

World Model-Guided Reinforcement Learning via Counterfactual User Engagement Simulation
Ang Li, Xin Xu, Bin Liang, Yue Ma, Fubang Zhao, Yangyang Kang and Kam-Fai Wong . . 45856

Elias in the Lighthouse, Again? Diagnosing Low Diversity in LLM Stories
Sil Hamilton and David Mimno 45877

TaRA: Training-Aware Low-Rank Adaptation Initialization
Taehyeon Kim and Eunhyeok Park 45885

UNISON: A Unified Sound Generation and Editing Framework via Deep LLM Fusion
Zhaoqing Li, Haoning Xu, Jingran Su, Yaofang Liu, Zhefan Rao, Huimeng Wang, Jiajun Deng, Tianzi Wang, Zengrui Jin, Rui Liu, Haoxuan Che and Xunying Liu 45902

Trait-Aware Policy Optimization for Autoregressive Multi-Trait Essay Scoring
Zhengyang Wang, Sanwoo Lee, Jiaxin Wang, Chenxi Miao, Weikang Li and Yunfang Wu . 45923

MOB-KV: Mixture-of-Basis for Key-Value Cache Compression
Jinseop Yeom, Taehyeon Kim and Eunhyeok Park 45939

Do LLM-Generated Web Apps Behave Correctly? A Contract-Grounded Benchmark for Behavioral Conformance
Chansung Park, Juyong Jiang, Jiasi Shen and Sunghun Kim 45959

SchemaToken: Tokenizing Database Schemas for Efficient and Robust Text-to-SQL Linking
Weibin Liao, Jingwen Liu, Rihong Qiu, Xin Gao, Xinyu Ma, Hongxin Ding, Yue Fang, Yifan Zhu, Yuheng Huang, Junfeng Zhao and Yasha Wang 45993

Agentic Time Machine as an Infrastructure for Future-Event Forecasting
Jingyi Chai, Bingyang Zheng, Xiangrui Liu, Hao Lu, Zihang Zhou, Tianchen Wang, Kemeng Zhang and Siheng Chen 46016

Investigating the Laws of Synonymy Evolution through a Fine-Grained Lemma-Based Framework
Eduardo Calò, Pierluigi Cassotti, Stefano De Pascale, Nina Tahmasebi and Pablo Mosteiro 46036

Fake News Court: A Multi-Agent Adversarial Framework for Robust Detection of LLM-Generated Fake News
Mengyao Peng, Hongyan Gu, Yiping Ma, Shiyu Hu and Xinli Huang 46053

LongWebBench: Evaluating Structural and Functional Webpage Generation in Long-Horizon Settings
Yi Zhao, Zhen Yang, Mengpan Chen, Mingde Xu, ShangHui Gong, Xijun Liu, Jibing Gong and Jie Tang 46064

RESKILL: Explicit Failure Attribution and Structured Repair for Interactive Language Agents
Mengyi Deng, Xin Li, Duyi Pan, Zilin Wang, Zhiwei Li, Zhijiang Guo and Wei Wang 46116

NE-R1: Enhancing Named Entity Recognition Model via Reinforcement Learning
Meixuan Chen, Hehan Li, Ruizhi Zhao, Xin Lu, Peizhi xu, Liwei Qian, LI Meifang, Shuanglong li, Hanmeng Liu, Xin Pei and Yanbiao Ma 46132

PRISM: Preference-aware Influence function based Data Selection Method for Efficient Fine-Tuning
Qihao Lin, Guanxu Chen, Dongrui Liu and Jing Shao 46150

Neuronal Redundancy Reduction and Collaborative Learning for Parameter-Efficient Mixture-of-Experts
Cheng Ding, Qin Chen, Jie Zhou, Hongjian Zhan and Liang He 46170

Format as Nuisance: Intervention-Inspired Cross-Format Distillation for Knowledge Conflicts
Yue Yu, Wang Luo, Shenghao Ye, Zepeng Bao, Jiacheng Liu, Yongqi Li, Wenxin Tang and Tieyun Qian 46190

Controlling and Assessing Appropriate Persona Use in LLM-based Dialogue Generation
Jongkyung Shin, Inkyu Lee and Chiehyeon Lim 46212

Parallelism Strategy Chaining for Fast Training Convergence
Minchul Kang, Changyong Shin, Younghun Go, Hyunho Lee, Jinwoo Jeong, Chuck Yoo and Gyeongsik Yang 46240

The Collaboration Tax: How Much LLM Multi-Agent Systems Pay to Coordinate
Weixiang Sun, Zehong Wang, Hong Huang, Colby Nelson, Yijun Ma and Yanfang Ye 46258

Beyond Prompts: Discovering Jailbreak Motifs in Sparse Feature Space
Zhongzhi Li, Jiahui Han, Jialing Tao, Huiqi Deng, Na Zou and Ninghao Liu 46282

Reasoning Depth and Environment Complexity: A Controlled Study of RLVR Data Allocation across Logical Reasoning Tasks
Yihua Zhu, Qianying Liu, Fei Cheng, Jiaxin Wang, Akiko Aizawa, Sadao Kurohashi and Hidetoshi Shimodaira 46316

Fast Block Diffusion Training via Multi-Mask Context Sharing
Tengyu Pan, Weilin Zhao, Yutao Sun, Yanghao Li, Jianyong Wang and Xu Han 46353

Will the User Ever Know? Covert Indirect Prompt Injection Attacks on Tool-Using LLM Agents
Yunseok Lee, Yunji Kim and Woojin Lee 46371

NimbleMoE: Collaboration-Aware Expert Merging with Sparse Compensation for Efficient Mixture-of-Experts LLMs
Zhiyuan Ning, Xiaoran Liu, Ruge Xu, Xinfei Guo, Ziwei He and Xipeng Qiu 46390

Systematic Failure Analysis of Open-Weight LLM Annotators for Classification Tasks: Taxonomy, Uncertainty Signals, and Deployment Boundaries
Nahid Hossain, Siam Rahman Karip, Md Faisal Kabir, Swakkhar Shatabda and Salekul Islam 46409

Communication Policy Evolution for Proactive LLM Agents
Xinbei Ma, Jiyang Qiu, Yao Yao, Zheng Wu, Yijie Lu, Xiangmou Qu, Jiaxin Yin, Xingyu Lou, Jun Wang, Weiwen Liu, Weinan Zhang, Zhuosheng Zhang and Hai Zhao 46467

Emergent Backdoors in Agentic Recursive Training
Xuan Wang, Yuliang Lu, Kejiang Ye, Cheng-zhong Xu and Xitong Gao 46496

ABLE: Representing and Mapping LLMs via Attribution-Based Large-model Embedding
Zirui Wang, Yusen Hou, Shaofeng Liang, Bowen Tian, Yanlin Zhang, Wenshuo Chen and Yutao Yue 46513

Boundary-Aware Expert Ordering Preservation with Next-Router Supervision for MoE Quantization
Hancheol Park, Geonho Lee and Tae-Ho Kim 46537

Can We Still Trace L1 Signals? Investigating the Resilience of Native Language Signals in the LLM Era
Nabelanita Utami and Ryohei Sasano 46549

Spatio-Temporal Audio Language Modeling for Dynamic Sound Sources
Oh Hyun-Bin, Kazuki Shimada, Yuhta Takida, Kim Sung-Bin, Toshimitsu Uesaka, Takashi Shibuya, Kyeongyoon Lee, Tae-Hyun Oh and Yuki Mitsufuji 46568

Counteraction-Aware Multi-Teacher On-Policy Distillation for General Capability Recovery with Domain Preservation
Tianlei Chen, Jiao Ou, Ziyuan Liu, Ruiming Tang, Jian Liang and Han Li 46591

Can Scientific Claims Be Removed from Large Language Models? A Systematic Evaluation of Claim-Level Unlearning
Snigdha Paul, Manasi Patwardhan and Arman Cohan 46608

The Correlation Mirage: Benchmark Dependence Collapses for Top-Performing LLMs
Raj Firke and Rajeswari Kannan 46629

SPRI: SVD-Partitioned Residual Initialization for Data-Constrained MoE Upcycling in Speech-to-Text Translation
Weiqiao Shan, Ruixiang Mao, Yuang Li, Yuhao Zhang, Yingfeng Luo, Tong Zheng, Chen Xu, Yucheng Qiao, Chunxiang Jin, Yi Yuan, Jingdong Chen, Tong Xiao and JingBo Zhu 46645

PCoMoE: Shifting MoE Inference from Monolithic Expert Selection to Fine-Grained Path Composition
Ziyan Gan, Fangxin Liu, Chenyang Guan, Junjie Wang, Ning Yang, Haomin Li, Xiang Li, Siran Yang, Jiamang Wang, Lin Qu, Zongwu Wang, Li Jiang and Haibing Guan 46662

Legible but Not Localized: Jailbreak Geometry in Mixture-of-Experts Language Models Is Attack-Specific
Yubin Im, Myeonghyun Kim, Seongmin Kim and Haehyun Cho 46673

Critical or Compliant? The Double-Edged Sword of Reasoning in Chain-of-Thought Explanations
Eunkyu Park, Wesley Hanwen Deng, Vasudha Varadarajan, Mingxi Yan, Gunhee Kim, Maarten Sap and Motahhare Eslami 46688

ReProF: Residual Prototype Fusion for Continual LLM Attribution
Yujie Wang, Zhendong Mao and Quan Wang 46709

Assembling the CREW: A Collaborative Multi-agent Reinforcement Learning Framework for Automated Related Work Generation
Hai-Dang Dang, Bao-Yen Pham, Bao Nguyen, Tran Thi Huong and Huynh Thi Thanh Binh 46725

MAPLE: Metadata Conditioned LLM Pretraining for Locale-Aware Question Answering
Anjishnu Mukherjee, Ziwei Zhu and Antonios Anastasopoulos 46746

Thinking Ahead: Prospection-Guided Retrieval of Memory with Language Models
Harshita Chopra, Krishna Kant Chintalapudi, Suman Nath, Ryen W White and Chirag Shah 46773

AdapTools: Adaptive Tool-based Indirect Prompt Injection Attacks on Agentic LLMs
Che Wang, Jiaming Zhang, Ziqi Zhang, Zijie Wang, Yinggui Wang, Jianbo Gao, Tao Wei, Zhong Chen and Wei Yang Bryan Lim 46795

Evaluating Coreference Consistency in Chinese-to-English Dialogue Translation
Xu Yang, Zhengxian Gong, Junhui Li, Jinlong Yang, Min Zhang, Shimin Tao and Daimeng Wei 46817

Beyond Shallow Alignment: How Post-Training Methods Determine Refusal Circuits And Steering Robustness
Hoang Cuong Nguyen, Mark Dras and Usman Naseem 46834

ATC: Agent Trajectory Control for Role-Contract Enforcement in Black-Box Multi-Agent Systems
Maisha Maliha, Vishal Pramanik and Sumit Kumar Jha 46861

Quantization Inflates Reasoning: Token Inflation as a Hidden Cost of Low-Bit Reasoning Models
Xinyu Lian, Walid Krichene, Beichen Huang, Masahiro Tanaka, Olatunji Ruwase, Li Zhang and Minjia Zhang 46880

Denoising the Future: Context-Aware Spectral Diffusion for Temporal Knowledge Graph Extrapolation
Yanglei Gan, Peng He, Run Lin, Peiyuan Jiang, Yifan Wang and Qiao Liu 46895

MACRO: A Benchmark for Clarification-Aware Multi-Agent Orchestration
Tao Li, Jianyi Chen, Yaqiang Wu, Jie Ma, Shangda Yang, Oguz H Elibol and Pradyumna S. Singh 46914

Double Trouble: Bilingual Pretraining Leaves Language-Conditioned Effects in Shared-Language Representations
Anjishnu Mukherjee, Ziwei Zhu and Antonios Anastasopoulos 46953

Show Me How You Cite: Citation Fidelity, Usage, and Topology in the Evidence Behavior of AI Scientific Writing
Jihwan Oh, Gyeonghun Sun, Jisoo Kim, Yuho Lee, Taewon Yun, Jisu Shin and Hwanjun Song 46978

Global Divergence, Local Convergence: Representation Geometry in SSMs and Transformers
Amit Ben-Artzy and Roy Schwartz 47005

</think> Doesn't Stop Reasoning: Analysis of Spurious CoT Termination
Seunghee Koh, Sungjae Choi, Minchan Kwon, Sunghyun Baek and Junmo Kim 47019

MVP: Multi-View-Guided Token Pruning for Vision-Language Models
Chenxi Wang, Dexing Zhong, Siyu Shi, Bowen Ping, Hang Yan, Huayang Li and Huikai Shao 47054

CLARUS: Hierarchical Hypergraph Modeling for Unsupervised Discharge Summarization
Gaetano Cimino, Mir Rayat Imtiaz Hossain, Laura De Santis, Amirhossein Abaskohi, Raymond T. Ng, Giuseppe Carenini and Vincenzo Deufemia 47070

Confident Rankings with Fewer Items: Adaptive LLM Evaluation with Continuous Scores
Esma Balkir, Alice Pernthaller, Marco Basaldella, Jose Hernandez-Orallo and Nigel Collier47092

SAFE: A Plug-and-Play Framework for Prompt Optimization in Text-to-Image Safety
Jisu Shin, Sol Lee, Sungrae Hong, A Young Kim and Mun Yong Yi 47109

RAT: RunAnyThing via Fully Automated Environment Configuration
Renhong Huang, Dongdong Hua, Yifei Sun, Sitao Ding, Hanyang Yuan, Daixin Wang and Yang Yang 47131

Generation-Gated Cascaded Routing for Large Language Models
Yanqiu Xu, Han Luo and Lei Feng 47162

Context-Aware Cluster Decoding: Semantic Anchor-Driven Coherence in dMLLMs
Yikai Zhao, Qiyan Zhao, Jiaquan Zhang, Xiaofeng Zhang, Xiaosong Yuan and Pengzhou Cheng 47182

Latent Bridges for Multi-Table Question Answering
Simone Varriale, Tamara Cucumides, Floris Geerts and Paolo Papotti 47201

COSE: Confidence-Orchestrated Self-Evolution for Effective LLM Reasoning
Bowen Wei, Nan Wang, Yuqing Zhou, Jinhao Pan and Ziwei Zhu 47229

Framework and Benchmark for Code-Driven Agentic Testing in Web Development
Bin Hong, Zhenchao Zhang, Jiyuan He, Kai Zhang and Zhenya Huang 47250

VLA-Trace: Diagnosing Vision-Language-Action Models through Representation and Behavior Tracing
Haoyuan Shi, Xiancong Ren, Yingji Zhang, Qinfan Zhang, Jiayu Hu, Haozhe Shan, Han Dong, Jinpeng Lu, Yinda Chen, Yi Zhang, Yong Dai and Xiaozhu Ju 47280

Not All Preferences Deserve Gradients: Understanding Gradient Utility in Offline Reasoning Alignment
Hui Wu, Hengyi Cai, Jinman Zhao, Xinran Chen, Ziheng Li, Zhejun Zhao, Shuaiqiang Wang, Yuchen Li and Dawei Yin 47316

OrchDebate: Adaptive Orchestration for Multi-Agent Debate via Self-Evolving Skills
Kaiyu Zhou, Yongsen Zheng, Xuanye Zhang and Kwok-Yan Lam 47331

Switch Attention: Towards Dynamic and Fine-grained Hybrid Transformers
Yusheng Zhao, Hourun Li, Bohan Wu, Yichun Yin, Lifeng Shang, Jingyang Yuan, Meng Zhang and Ming Zhang 47351

Beyond Chunking: Efficient Global Pooling for Holistic Long-Document Representation
Hoin Jung and Xiaoqian Wang 47372

How to Utilize Complementary Vision-Text Information for 2D Structure Understanding
Jiancheng Dong, Pengyue Jia, Derong Xu, Jiawei Cheng, Jingyu Peng, Chao Zhang, Bowen Liu, Xin Sun, Lixin Su, Shuaiqiang Wang, Dawei Yin and Xiangyu Zhao 47398

RiskCueBench: Evaluating Vision Language Models for Anticipatory Risk Reasoning in Videos
Sha Luo, Yogesh Prabhu, Timothy Ossowski, Kaiping Chen and Junjie Hu 47417

Where Does Robustness Live? Neuron-Guided Adaptation for Retrieval-Augmented Language Models
Jae O Lee, Jaemin Kim, Sumyeong Ahn and Seo Yeon Park 47450

You Know What I Mean: A Benchmark for Agentic Conversational Reference Grounding
Karen Fuchs, Uri Katz and Yoav Goldberg 47470

Depth-Attention: Cross-Layer Value Mixing for Language Models
Boyi Zeng, Yiqin Hao, Zitong Wang, Shixiang Song, He Li, Feichen Song, Yifan Liu, Ziwei He, Xinbing Wang and Zhouhan Lin 47493

Harnessing LLM Agents with Skill Programs
Hongjun Liu, Yifei Ming, Shafiq Joty and Chen Zhao 47515

Prompts Without Evidence: How Neuroimaging Mentions Shift Clinical Vision-Language Model Predictions
Doan Nam Long Vu and Simone Balloccu 47550

Learning to Better Exploit Visual Evidence for Multimodal Entity Linking
Yingyao Ma, Han Wang, Yuning Wu, Yifan Xue, Jiasong Wu, Lotfi Senhadji and Huazhong Shu 47579

Distill Where You Fail: Recovering Learning Signals of Negative RL-Groups from Adaptive Teacher Guidance
Zhuowen Han, Jinwei Xiao, Zhengxi Lu, Renren Jin, Zhiyuan Yao, Yuxin Liu, Hongyan Hao, Yueqing Sun, Yu Yang, Xierui Song, Xunliang Cai and Deyi Xiong 47593

Beyond Max Tokens: Stealthy Resource Amplification via Tool Calling Chains in LLM Agents
Kaiyu Zhou, Yongsen Zheng, Yicheng He, Meng Xue, Xueluan Gong, Yuji Wang, Xuanye Zhang and Kwok-Yan Lam 47607

MigrationNarrate: A Dataset for Detection of Migration Narratives in YouTube Videos
Fatima Haouari, Carolina Scarton and Kalina Bontcheva 47624

When Models Edit Too Much: On the Fidelity of Minimal Code Edits
Tongyao Zhu, Lim Wei Hern and Min-Yen Kan 47646

Out of Sight, Not Out of Mind: Unveiling Latent Attack in Latent-based Multi-Agent Systems
Chenxi Wang, Ruiyang Huang, Jiayan Sun, Lei Wei and Yifan Wu 47668

DA-DLM: Explicitly Modeling Token Dependencies in Diffusion Language Models
Pengyu Ji, Zichen Zhang, Xiang Hu and Kewei Tu 47697

GSAR: Goal-State-Anchor Rewards for Mobile GUI Agents with Self-Evolving Data Synthesis
Long Zhang, Yuhan Chen, Chaoran Zhang, Wanxia Cao, Kun Huang, Pengzhi Gao, Wei Liu, Jian Luan, Chenliang Li and Lixin Zou 47714

Surprisal Theory is Tautological (without Rational Grounding)
Ryan Cotterell 47731

The Geometry of Low-Resource Language Representations
Francois Meyer and Jan Buys 47748

Boosting Text-to-Chart Retrieval through Synthesized Hierarchical Insights
Yifan Wu, Lutao Yan, Yizhang Zhu, Yenchi Tseng, Yong Wang, Jiannan Wang, Nan Tang and Yuyu Luo 47764

Who Wins the Conflict? Mechanistic Interpretability of Text Bias in Audio LLMs
Hyebin Cho, Suho Yoo, Jaehyuk Jang, Changick Kim and Joon Son Chung 47781

TAME: A Trustworthy Test-Time Evolution of Agent Memory with Systematic Benchmarking
Yu Cheng, Yongkang Hu, Jiuan Zhou, Yushuo Zhang, Yihang Chen, Huichi Zhou, Mingang Chen, Zhizhong Zhang, Kun Shao, Yuan Xie and Zhaoxia Yin 47800

PA3: Policy-Aware Agent Alignment through Chain-of-Thought
Shubhashis Roy Dipta, Daniel Bis, Kun Zhou, Lichao Wang, Benjamin Z. Yao, Chenlei Guo and Ruhi Sarikaya 47819

Entropy-Regularized Rank-Masked Policy Optimization for Test-Time Reinforcement Learning in Code Generation
Jiacheng Xu, Feng Chen, Xiuneng Xu and Bo An 47843

EpiQAL: Benchmarking Large Language Models in Epidemiological Question Answering and Reasoning
Mingyang Wei, Dehai Min, Zewen Liu, Yuzhang Xie, Guanchen Wu, Ziyang Zhang, Carl Yang, Max SY Lau, Qi He, Lu Cheng and Wei Jin 47858

Uncertainty-Aware Hybrid Retrieval for Long-Document RAG
Hoin Jung and Xiaoqian Wang 47893

OSPD: On-Policy Self-Distillation for Persona-Consistent Dialogue
Rui Xu, Yikai Zhang, Aili Chen, Zicheng Zhao, Xu Yinghui and Libo Wu 47912

Perspectives on Cross-Lingual Consistency in LLMs for Medical Questions
Minh Duc Bui, Mario Sanz-Guerrero, Abteen Ebrahimi, Sagi Shaier, Peter Herbert Kann, Manuel Mager and Katharina von der Wense 47934

Privacy-Preserving Generation of Clinical Narratives from Medical Terminologies
Yuping Wu, Viktor Schlegel, Warren Del-Pinto, Srinivasan Nandakumar, Iqra Zahid, Yidan Sun, Hao Li, Usama Farghaly Omar, Amirah Jasmine, Arun-Kumar Kaliya-Perumal, Chun Shen Tham, Gabriel Connors, Anil Anthony Bharath and Goran Nenadic 47949

EviNoteRAG: Enhancing RAG Models via Quality-Augmented Evidence Notes
Yuqin Dai, Guoqing Wang, Yuan Wang, Zhouheng Yao, Kairan Dou, Kaichen Zhou, Zhanwei Zhang, Shuo Yang, Fei Tang, Jun Yin, Pengyu Zeng, Can YI, Changhua Meng, Yuchen Zhou, Yongliang Shen and Shuai Lu 47970

Knowledge Boundaries Are Not Defined by Prompts : Improving the Reachability of Knowledge through Question-Level Probing
Wang Bin, Yilu Zhong, Shuping Chen and Hui LI 47991

Divide by Question, Conquer by Agent: SPLIT-RAG with Question-driven graph partitioning
Ruiyi Yang, Hao Xue, Imran Razzak, Shirui Pan, Hakim Hacid and Flora D. Salim 48009

One Trace Is Not Enough: Paired Probing for Runtime Backdoor Detection in Limited-Black-Box LLMs
Yansu Wang, Feiyan Chen, Rui Zhang and Kaizhu Huang 48027

Natural-Language Temporal Grounding in Hour-Long Videos is a Search Problem: A Benchmark and Empirical Decomposition
Sukmin Seo and Geewook Kim 48044

The Language–Energy Divide: Measuring Energy Costs of Multilingual LLM Inference
Naihao Deng, Alissa Shen, Yiming Feng, Joan Nwatu, Jae-Won Chung, Mosharaf Chowdhury, Yulong Chen and Rada Mihalcea 48057

PEAR: Permutation-Equivariant Adaptive Routing Multi-Agent Debate
Yang Feng, Ziwei Xu, Xia Hu and Fengxiang He 48076

ARM: Role-Conditioned Neuron Transplantation for Training-Free Generalist LLM Agent Merging
Zhuoka Feng, Kang Chen, Sihan Zhao, Kai Xiong, Yaoning Wang, Yu Minshen, Junjie Nian, Changyi Xiao, Yixin Cao and Yu-Gang Jiang 48104

MulFeRL: Enhancing Reinforcement Learning with Verbal Feedback in a Multi-turn Loop
Xuancheng Li, Haitao Li, Yujia Zhou, Yiqun Liu, Min Zhang and Qingyao Ai 48119

Re^{3}Cap: Retrieval-Guided Refinement for Image Captioning Enhancement via Reinforcement Learning
Haonan Jia, Shichao Dong, Zenghui Sun, Jiawen Zheng, Ziqi Miao, Gege Shi, Qiuyu Zhao, Jinsong Lan, Xiaoyong Zhu and Bo Zheng 48154

Different Demographic Cues Yield Inconsistent Conclusions About LLM Personalization and Bias
Manuel Tonneau, Neil Sehgal, Niyati Malhotra, Sharif Kazemi, Victor Orozco-Olvera, Ana Maria Munoz Boudet, Lakshmi Subramanian, Samuel Fraiberger, Sharath Chandra Guntuku and Valentin Hofmann 48170

Beyond Memorization: Distinguishing Between Pattern-Based and Epistemic Reasoning in LLMs Using Epistemic Puzzles
Adi Gabay, Gabriel Stanovsky and Liat Peterfreund 48199

HuggingR^4: A Progressive Reasoning Framework for Discovering Optimal Model Companions
Shaoyin Ma, Chenggong Hu, Huiqiong Wang, Li Sun, Mingli Song and Jie Song 48219

LexAgentHallu: A Hierarchical Benchmark for Profiling Hallucinations in Legal Agents
Yujin Zhou, Mingxuan Zheng, Chuxue Cao, Huang Yidan, Jiale Chen, Yike Guo and Sirui Han 48241

Generating Reports or Repeating Templates? Measuring and Mitigating Template Collapse in 3D CT Report Generation
Tom Maye-Lasserre, Yitong Li, Bailiang Jian, Morteza Ghahremani, Benedikt Wiestler and Christian Wachinger 48270

Is Position Bias in Dense Retrievers Built In—or Learned from Data?
Daegon Yu, SeungYoon Han and Woomyoung Park 48287

Flip, Don't Shuffle: Watermarking LLMs at the Speed of Inference
Simone Ceppi and Ignacio Sanchez 48312

One Example Is Enough to Pass Fairness Benchmarks: Rethinking Fairness Evaluation for Aligned LLMs
Naihao Deng, Samee Arif, Shuaichen Chang, Yulong Chen and Rada Mihalcea 48329

SCoNE : Selective Context-aware Neuron Editing for Robust Retrieval-Augmented Generation
Chaewon Kim and Seo Yeon Park 48352

Language Proficiency Assessment from Eye Movements in Naturalistic Passage Reading
Shachar Frenkel, Ido Falah, Omer Shubi and Yevgeni Berzak 48364

How Much Can We Trust LLM Search Agents? Measuring Endorsement Vulnerability to Web Content Manipulation
Yimeng Chen, Zhe Ren, Firas Laakom, Yu Li, Dandan Guo and Jürgen Schmidhuber 48390

PowerOPD: Stabilizing On-Policy Distillation with Bounded Power Transformation
Anhao Zhao, Junlong Tong, Yingqi Fan, Ping Nie, Wenjie Li and Xiaoyu Shen 48414

COIN: A Live, Large-scale Benchmark for Semantics-to-Input Code Reasoning
Jiho Kim, Cen Zhang, Youngjoon Kim, Dongkwan Kim, Brian Junekyu Lee, Dae R. Jeong and Taesoo Kim 48428

Low-Resource Preference Adaptation of LLMs via Activation-Based Label Propagation
Alessio Galatolo and Meriem Beloucif 48454

Subtraction Gets You More: Gap-Aware Retrieval for Multimodal Multi-Hop QA
Sunah O and Jay-Yoon Lee 48472

Understanding & Isolating the Assistant State in Instruction-Tuned Models
Aishwarya Maheswaran and Maunendra Sankar Desarkar 48491

Enhancing Repository-Level Code Completion with Reinforcement Learning in Real-World Scenarios
Zhenlong Dai, Zhuoluo Zhao, Xiu Tang, Sai Wu, Chang Yao, Zhipeng Gao and Jingyuan Chen 48512

Vista: An Efficient Late Interaction Engine for Multi-Vector Visual Document Retrieval
Haolin Jiang, Jinwen Pan, Abhishek Saini, Weihang Li and Hang Liu 48529

GraphProfiler: Source-Linked Sensitive Attribute Inference via Personal Knowledge Graphs
Ahmed Sohair Khan, Estrid He, Chenglong Ma, Monica Wachowicz and Elham Naghizade 48544

ConFormBench: Evaluating Evidence-Supported Record Maintenance in Conversational Form Filling
Timothy Veigel and Sven Mayer 48575

R2MED: A Benchmark for Reasoning-Driven Medical Retrieval
Xiangxu Zhang, Lei Li, Zheng Liu and Xiao Zhou 48600

When Tokenizers Fail: Byte-Level Chunking for Zero-Shot Transfer to Low-Resource Languages
Sanjeev Kumar, Atsuki Yamaguchi and Nikolaos Aletras 48638

Beyond Final Success: Semantic Progress Diagnostics for Agent Trajectory Evaluation
Nan Xue, Tinghao Wang and Li Tang 48652

Beyond Semantic Accuracy: Consequence-Aware Evaluation for Safety-Critical Language Understanding
Yujing Chang, Duc-thinh Pham, Van-Phat Thai, Chunyao Ma, Yash Guleria, Pham Nhut Huy and Sameer Alam 48673

Give it Space! Explicit Disentangling of Positional and Semantic Representations in Encoders
Pierre-Antoine Lequeu, Camille Barboule and Benjamin Piwowarski 48709

Group-Aware Adaptive Retrieval for Evidence Navigation
June Park, Jun Kwon, Jonghyo Kim and Jongwuk Lee 48728

Task-Oriented Moment Localization for Long Video Understanding with Near-Second Coverage
Xiangrui Liu, Minghao Qin, Yan Shu, Zhengyang Liang, Yang Tian, Chen Jason Zhang, Bo Zhao and Zheng Liu 48748

RASET: Router-Agnostic Safety-Critical Expert Tuning Exposes Localized Safety Enforcement Failures in Mixture-of-Experts LLMs
Zhibo Zhang, Yuxi Li, Zhen Ouyang, Ling Shi and Kailong Wang 48770

Pivot-SD: Efficient Self-Distillation for Masked Diffusion Language Models
Seo Hyun Kim, Sunwoo Hong, Younwoo Choi, Chen-Hao Chao, Se-Young Yun and Rahul G Krishnan 48790

Steer on Rubrics: Auditable Steering of Schwartz Values in Language Models
Zichen Xu, Geng Zhao, Aoxiang Qin, Li Zhou and Benyou Wang 48807

Warm Up with a Rubric, Then Merge: Decoupling Grounding Gains from Reasoning Costs in Vision-Language Models
Jingfan Chen, Lin Wang, Jingran Su and Li Qing 48839

STORM: Stepwise Token Optimization with Reward-Guided Beam Search
Arthur Satouf, Giulio D'Erasmo, Yuxuan Zong, Habiboulaye Amadou Boubacar, Pablo Piantanida and Benjamin Piwowarski 48856

LEAP: Learning Edit-Adaptable Parameters for Multi-Hop Knowledge Propagation
Taichi Shiba, Max Müller-Eberstein, Masaru Isonuma, Yohei Oseki and Yusuke Miyao . . . 48872

Reinforcement Learning Can Amplify Emergent Misalignment from Harmless Rewards
Magnus Jørgenvåg, David Kaczér, Lasse Ruttert, Marvin Gülhan, Lucie Flek and Florian Mai 48888

LLM Explainability with Counterfactual Chains and Causal Graphs
Nirit Nussbaum-Hoffer, Nitay Calderon, Liat Ein-Dor and Roi Reichart 48927

Language Shapes Instruction Hierarchy Compliance in Multilingual LLMs
Jiwon Moon, Yerin Hwang and Kyomin Jung 48968

Do LLMs Understand Backend Data-Layer Semantics? A Controlled Study of Cross-Endpoint Issues
Francesco Dente, Dario Satriani, Donatello Santoro, Enzo Veltri and Paolo Papotti 49005

SCORE: Similarity-Aware Contextual Overlap-Redundancy Eviction for Efficient KV Cache Compression in LLMs
Gilha Lee, Seungil Lee and Hyun Kim 49020

How You Ask Matters! Adaptive RAG Robustness to Query Variations
Yunah Jang, Megha Sundriyal, Kyomin Jung and Meeyoung Cha 49039

OmniReasoner: Thinking with Long Audio-Video via Native Tool Use
Yu Chen, Qiufeng Wang, Caorui Li, Ziyu Xiong, Yi-Dong Wang, Mingqi Gao, Shuman Liu, Biao Liu, Chunfeng Yang, Anxiang Zeng, Haibo Zhang and Chaofan Chen 49078

Defending Agentic Fact-Checking Against Adversarial Attacks
Fanzhen Liu, Tingyang Xu, Xixun Lin, Haolong Xiang, Ge Zhang, Anan Du and Shuchao Pang 49097

Steering Language Models Before They Speak: Logit-Level Interventions
Hyeseon An, Shinwoo Park, Hyundong Jin and Yo-Sub Han 49123

LoopMoE: Unifying Iterative Computation with Mixture-of-Experts for Language Modeling
Wenkai Chen, Tianshu Li, Wenyong Huang, Yichun Yin, Lifeng Shang and Chengwei Qin 49144

SV-Detect: AI-generated Text Detection with Steering Vectors
Mikhail Vishnyakov and Tatiana Gaintseva 49160

Beyond Homogeneous Attention: Memory-Efficient LLMs via Fourier-Approximated KV Cache
Qiqi Wang, Siyang He, Xiaoran Liu, Ruixiao Li, Yuerong Song, Zhigeng Liu, Mianqiu Huang, Zengfeng Huang, Qipeng Guo, Ziwei He and Xipeng Qiu 49187

MolSC: Leveraging Substituent Contributions to Enhance Fine-grained Molecular Understanding in LLMs
Hyuntae Park, Sooyeon Kim, Jiwon Park and SangKeun Lee 49206

Too Many Cooks Spoil The Broth: Auditing Multi-Agent LLM Councils For Medical Reasoning
Mukul Joshi, Pratik Jalan, Akhilesh Magotra and Kshitij Sharad Jadhav 49230

T2ER: A Chinese Benchmark for Text-to-Executable-Rule Generation in Medical Insurance
Hanjing Yan, Rubin Li, Yuxuan Lin, Siyi Zhu, Yuxiang Wang, Guangya Yu, Yuxiang Chu, Yongqi Fan, Yinan Wu, Bao Wei, Weiyan Zhang and Tong Ruan 49246

Improving Latent Reasoning in LLMs via Soft Concept Mixing
Kang Wang, Xiaocheng Luo, Xiangyu Duan, Yu Yao, Zaifu Zhan and Min Zhang 49265

World-Model Evolving Agents
Leonardo Ranaldi 49279

Abstract Activation Spaces for Content-Invariant Reasoning in Large Language Models
Gabriele Maraia, Marco Valentino, Fabio Massimo Zanzotto and Leonardo Ranaldi 49295

What Transfers Across Cuneiform? Script-Level Substrate and Language-Specific Regularities via Segmentation and Representation
Chuanjun Zhou and Adam G Anderson 49324

LangMAP: A Language-Adaptive Approach to Tokenization
Clara Meister, Suchir Salhan, Andrzej Szablewski, Pietro Lesci, Paula Buttery and Tiago Pimentel 49341

E2Rank: Unifying Text Embedding and Listwise Reranking for Effective and Efficient Search
Qi Liu, Yanzhao Zhang, Mingxin Li, Dingkun Long, Pengjun Xie and Jiaxin Mao 49363

MACBench: Benchmarking Multi-turn Agentic Coding with Evolving Requirements in Frontend Web Development
Jiayi Gui, Bosi Wen, Haoke Zhang, Pei Ke, Yukuo Cen, Dayong Yang, Minlie Huang, Jie Tang and Cunxiang Wang 49386

VISPA: Pluralistic Alignment via Automatic Value Selection and Activation
Shenyan Zheng, Jiayou Zhong, Anudeex Shetty, Heng Ji, Preslav Nakov and Usman Naseem 49402

IterSynth: Rethinking Deep Search Agents via Role-Decoupled Iterative Synthesis
Xingyu Wu, Yuchen Yan, Zhengxi Lu, Siqi Chen, Xin Zhang, Aiting Liu, Chao Deng, Jie Liu, Jin Ma, Jian Shao, Jun Xiao and Yongliang Shen 49432

Language Identification for Non-Standard Varieties: Language Varieties as Geolects on a Geodesic Grid
Sina Ahmadi and Rico Sennrich 49459

On the Efficiency-Safety Dilemma in Large Reasoning Models
Yifei Yang, Zouying Cao, Xingrui Wang, Xiao Zhou, Yuexian Li, Dongjie Yang and Hai Zhao 49482

Shortcut-Controlled Evaluation for AI-Generated Text Detection
Hee-Young Jung, Jae-Eon Ahn, Seong-Jin Park and Kang-Min Kim 49501

Does Modern Standard Arabic (MSA) Dominate Arabic Dialects in LLMs? A Representation-Level Analysis
Abdu Sallouh, Nicholas Popovič and Michael Färber 49531

MessageSieve: Runtime Message Pruning for Accurate and Efficient LLM-Based Multi-Agent Communication
Nianhao You, Yubo Dong, Jun Wang, Hehe Fan and Yi Yang 49544

UniComp: A Unified Evaluation of Large Language Model Compression via Pruning, Quantization, and Distillation
Jonathan von Rad, Yong Cao and Andreas Geiger 49567

A retrospective analysis on the use of LLMs to study infant syntax learning
Hélie Bazin, Anouk Barberousse and François Yvon 49590

FreqBLiMP: Frequency-Controlled Minimal Pairs Reveal Robustness and Fragility of LLMs Under Lexical Rarity
Tyrone White and Yuki Arase 49607

Self-Correcting Agentic RAG via Memory-Grounded Failure Localization
Haoyang Zhong, Yifei Sun, Shiwei Liu, Jie li, Chunping Wang and Yang Yang 49629

WearerText: Benchmarking Wearer-Centered Scene Text Understanding in AI-Glasses Videos
Yaping Zhang, Mei Tu, Jinting Wu, Qixuan Zhang, Xiaoyu Chen, Mengchao Zhang, Yuchen Huang, Junshen Chen, Fan Zhang, Lu Xiang, Junnan Zhu, Yang Zhao, Yu Zhou and Chengqing Zong 49653

MultiSynt/MT: Trillion-Token Multi-Parallel Pretraining Data Translated Across 36 Languages
Maximilian Idahl, Jörg Tiedemann, Sampo Pyysalo, David Salinas, Tomasz Galica, Shenbin Qian, Tudor Nicolae Mateiu, Zihao Li, Anna Lokrantz, Fedor Vitiugin, Andre Martins, Jenna Kanerva, Filip Ginter, Matthias Lindemann, Tim Isbister, Birger Moell, Jonas Lindh, Jan Hajič, Jenia Jitsev, Andrey Kutuzov, Stephan Oepen and Gema Ramírez-Sánchez 49670

Opacity Is Not Privacy: Information-Constrained Latent Communication for Multi-Agent LLM Systems
Yao Gong, Long Xiang, Di Liang, Peiyang Liu, Zekai Lin, Mengze Hong, Xianjie Wu and Minlong Peng 49695

CForce: Boosting Parallel Decoding for dLLMs via Consistency Forcing
Yuji Ren, Chenkai Xu, Zhuocheng Gong, Jianguo Li and Zhijie Deng 49718

Unified Context Evolution for LLM Agents
Zixuan Zhu, Yitong Hu, Yong Dai, Junfeng Fang, Chunyang Jiang, Senkang Hu and Yuzhi Zhao 49735

Exploring Collaboration between a language and a non-language agent
Harini S I, Somesh Kumar Singh, Yaman Kumar Singla, Rajiv Ratn Shah, David Doermann and Balaji Krishnamurthy 49764

Steering from Random Quotients: A Sample-Complexity Theory of Polysemantic Contrast Pairs
Khanh Nguyen Quoc, Luong Doan, Duc Hai Nguyen, Mai Phan Quoc Hung, Ngoc Mai Vu, Phong Ho, Nhung Duong and Tuan Do 49800

CultureVidBench: Benchmarking Cultural Understanding in Text-to-Video Generation
Xianjing Han, Yuhan SU, Yang Deng, Dong Ma, Wee Peng Tay and Bin Zhu 49814

Can LLM-as-a-Judge Reliably Verify Rubrics in Agentic Scenarios?
Yangda Peng, Yunjia Qi, Haotian Xia, Guanzhong He, Xintong Shi, Richeng Xuan, Songyuanyi Lu, Yixian Liu, Zhichao Hu, Lliu Yuhong and Hao Peng 49833

ABED: Assumption-Based Bidirectional Entailment Diagnosis for Open-World Logical Reasoning
Seyeon Jeong, Sojeong Park and Beakcheol Jang 49852

SAFER: Spectral Correlation Prior-based Attention Signature Flow for Enhanced Robustness in Harmful Prompt Detection
Neeva Oza, Kolavennu Mani Kanta Krishna Kumar and Rahul Mishra 49877

Differentiable Geometric Alignment for LLM Spatial Reasoning: Training-Time Symbolic Constraints, Solver-Free Inference
Yuhao Liao, Zijian Zeng, Huiming Yang and Dingding 49905

SpeechRole: A Large-Scale Dataset and Benchmark for Evaluating Speech Role-Playing Agents
Changhao Jiang, Jiajun Sun, Yifei Cao, Jiabao Zhuang, Xinmeng Che, Hui Li, Xiaoran Fan, Ming Zhang, Junjie Ye, Shihan Dou, Zhiheng Xi, Jingqi Tong, Yilong Wu, Baoyu Fan, Tao Ji, Tao Gui, Qi Zhang and Xuanjing Huang 49929

Beyond Linearization: Attributed Table Graphs for Table Reasoning
Yuxiang Wang, Junhao Gan, Shengxiang Gao, Shenghao Ye, Zhengyi Yang and Jianzhong Qi 49955

CopyShield: A Cross-Level Benchmark of Copyright Defenses in LLMs
Maryam Alshehyari, Dushyant Singh Chauhan, Samuele Poppi, Martin Takáč, Salem Lahlou and Nils Lukas 49975

GRKV: Global Regression for Training-Free KV Cache Compression in Long-Context LLMs
Junjie Peng, You Wu, Haoyi Wu, Jialong Han, Xiaohua Xie, Kewei Tu and Jianhuang Lai . 49990

A Universal Vibe? Finding and Controlling Language-Agnostic Informal Register with SAEs
Uri Z. Kialy, Avraham Shtarkberg and Ayal Klein 50015

Code Agents Don't Understand Root Cause:Evaluating Code Agents on Linux Kernel Fault Localization
Yuekeng Li, Ao Xu, Jie Shi, Sihang Jiang, Lipeng Ma, Bo Xu and Yanghua Xiao 50033

Training-Free Generative Retrieval
Yingchen Zhang, Ruqing Zhang, Jiafeng Guo, Maarten de Rijke, Fuyu Lv and Xueqi Cheng50054

SiMing-Bench: Evaluating Procedural Correctness from Continuous Interactions in Clinical Skill Videos
Xiyang Huang, Jiawei Lin, Keying Wu, Jiaxin Huang, Kailai Yang, Renxiong Wei, Cheng Zeng, Jiayi Xiang, Ziyan Kuang, Min Peng, Qianqian Xie and Sophia Ananiadou 50070

ESSA: Evolutionary Strategies for Scalable Alignment
Daria Korotyshova, Boris Shaposhnikov, Alexey Malakhov, Alexey Khokhulin, Nikita Surnachev, Kirill Ovcharenko, George Bredis, Alexey Gorbatovski, Viacheslav Sinii and Daniil Gavrilov . . . 50104

QuaNAG: LLM-driven Memory-bounded Neural Architecture Generation via Quantization-branched Exploration
Yingqi Peng, Yifan Lin, Kaijie Gong, Yi Gao and Wei Dong 50143

Implicit Unwatermarking in Pseudo-Reference Reranking for LLM Watermarking
Sungwoo Han, Jingun Kwon, Hidetaka Kamigaito and Manabu Okumura 50175

A Survey on Natural Language User Profiles for Recommendation: Methods, Datasets, and Metrics
Mariam Arustashvili and Krisztian Balog 50188

Beyond Survival: Evaluating LLMs in Social Deduction Games with Human-Aligned Strategies
Zirui Song, Yuan Huang, Junchang Liu, Haozhe Luo, Chenxi Wang, Lang Gao, Zixiang Xu, Bairui Zhang, Mingfei Han, Xiaojun Chang and Xiuying Chen 50207

Justification of Emotional Appeal in Computational Argumentation
Yanran Chen and Steffen Eger 50244

HEAR-RAG: A Hypergraph-Based Evidence-Aware Retrieval-Augmented Generation Framework for Multi-Hop Question Answering
Changbin Lv, Weihua Wang, Lianlong Tao, Jiale Xing and Feilong Bao 50274

Decouple to Forget: Semantic Mediation for Sensitive Unlearning in Large Language Models
Jingyue Cong, Andy Song, Yang Yu, Tabinda Sarwar, Xinshu Li, Ziqi Xu, Xun Yi, Ke Wang, Bo Han and Estrid He 50290

Multi-Field Tool Retrieval
Yichen Tang, Weihang Su, Min Zhang, Yiqun Liu and Qingyao Ai 50316

Span-Level Machine Translation Meta-Evaluation
Stefano Perrella, Eric Morales Agostinho and Hugo Zaragoza 50342

Causal Abstention for Cost-Aware Language-Agent Collaboration
Lei Wei, Ruiyang Huang, Zhaolu Kang, Qi Liu, Xiao Peng, Haoran Liu, Zonghao Ying, Haoyu Wang, Ye Shen, Xu Dong, Hongyu Li, Kesu Wang, Yukun Lian and Sicong Xie 50362

AuthorMix: Modular Authorship Style Transfer via Layer-wise Adapter Mixing
Sarubi Thillainathan, Ji-Ung Lee, Michael Sullivan and Alexander Koller 50382

FineWeb-CLaR: Culture, Language, and Region Annotations for Benchmark-Aligned Corpus Auditing
Yusser Al Ghussin, Eva Gavaller, Cristina España-Bonet, Josef Van Genabith and Simon Ostermann 50417

Why Users Go There: World Knowledge-Augmented Generative Next POI Recommendation
Qiuyu Ding, Heng-Da Xu, Wei Zhang, Dongyi Lv, Changda Xia, Feng Xiong and Mu Xu . 50446

GraphRAG-Router: Learning Cost-Efficient Routing over GraphRAGs and LLMs with Reinforcement Learning
Dongzhe Fan, Chuanhao Ji, Zimu Wang, Tong Chen and Qiaoyu Tan 50460

Negation Without Confusion: Decomposing Bias and Recomposing Style in Text Embeddings for Negation-Aware Image Generation
Sooyeon Park, Hemraj Singh, Kyunghyun Min, Youngjae Yu and Sung-Bae Cho 50492

CMIE: Cross-Modal Interaction Erasure for MLLM Safety Unlearning via Counterfactual Localization and Subspace Nullification
Jiatian Cao, Linyi Cai, Mengqi Wang, Jinduo Liu, Junzhong Ji, Yongping Du and Honggui Han 50518

MemWM: Memory-Augmented Text-Based World Model
Yujun Wang, Tao Zhang, Jinhe Bi, Aniri, Wenxuan Ye, Boliang Liu, Sikuan Yan, Shuning Wang, Xuebing Zhou, Soren Pirk, Hinrich Schuetze and Yunpu Ma 50542

Tracing Multilingual Representations in LLMs with Cross-Layer Transcoders
Abir Harrasse, Florent Draye, Punya Syon Pandey, Zhijing Jin and Bernhard Schölkopf . . . 50565

MobQA: A Benchmark Dataset for Semantic Understanding of Human Mobility Data through Question Answering
Hikaru Asano, Hiroki Ouchi, Akira Kasuga and Ryo Yonetani 50602

Adaptive Reasoning for Efficient Generative Retrieval
Yingchen Zhang, Ruqing Zhang, Jiafeng Guo, Wenjun Peng, Sen Li, Fuyu Lv and Xueqi Cheng 50634

What Matters for Aggressive Decoding-Time KV Eviction? Temporal Aggregation and Ranking Preservation
Bo Zeng, Yu Zhao, Yefeng Liu, Zhihong Lu, Xuanfan Ni and Xintong Wang 50648

Grounding LLM Agent Intent in Multi-Module Environments with Executable Action Programs
Jun Zhang, Boyuan Sun, Jinghua Piao, Haoyu Huang, Depeng Jin and Yong Li 50666

MERIT: Matching Expertise via Rubric-Informed Training for Reviewer Assignment
Zixuan Yang, Yibo Zhao, Weicong Liu and Xiang Li 50686

Is Memorization Context-Sensitive? Prefix-Based Extraction Beyond Isolated Prefixes
Ali Satvaty, Narjes Sharafi, Jirui Qi, Suzan Verberne and Fatih Turkmen 50711

ArtifactsBench: Bridging the Visual-Interactive Gap in LLM Code Generation Evaluation
Chenchen Zhang, Yuhang Li, Jiaheng Liu, Can Xu, Ao Liu, Changzhi Zhou, Ken Deng, Dengpeng Wu, Guanhua Huang, Kejiao Li, Qi Yi, Ruilin Lv, Shihui Hu, Yue Zhang, Yuhao Jiang, Zenan Xu, Yuanxing Zhang, Wiggin Zhou, Bo Zhou and Fengzong Lian 50735

MM-VeriRubric: Verifiable Evidence-Grounded Rubrics for Multimodal Reward Models
Zhiyong Cao, Bo Wang, Jintao Li, Shuxia Liu, Kairan Zhang, Aijia Wang, Yi-Liang Wei and Sicheng Zhou 50766

Entity tracking emerges in sub-billion parameter language models and exceeds human performance in naturalistic narratives
Karolina Drożdż and Micha Heilbron 50787

Evaluating Time Awareness and Cross-modal Active Perception of Large Models via 4D Escape Room Task
Yurui Dong, Ziyue Wang, Dairu Liu, Shuyun Lu, Xuechen Liu, Fuwen Luo, Peng Li and Yang Liu 50801

Athena: Vulnerability-Affected Library Identification via Knowledge Graph Completion
Phong Trinh Duy, Trang Dang Yen, Hung Nguyen-Huu, Huynh Quyet Thang, Bach Le, Dieu Hoang Vu, David Lo and Thanh Le-Cong 50831

Surfacing Program Structure with Auxiliary Signals for Low-Resource Code Translation
Federico Ranaldi, Nickil Maveli, Shay B. Cohen and Fabio Massimo Zanzotto 50850

Cross-Centaur: Modeling Individual Decision-Making Across Psychological Tasks with Language Models
Hikaru Asano, Hidezo Suganuma, Ryutaro Mori and Yukino Baba 50868

INTRYGUE: Induction-Aware Entropy Gating for Reliable RAG Uncertainty Estimation
Alexandra Kuleshova, Andrei Volodichev, Daria Kotova and Alexey Zaytsev 50891

Accelerated Drift of Scientific Prose in the LLM Era: Evidence from an Interrupted Time-Series Analysis
Hanna Hubarava and Benjamin Victor Ineichen 50910

WebSync: A Self-Evolving Web Agent Framework for Co-Evolving Strategic Advice and Executable Skills
Yuanlei Wang, Haohao Luo, Zihao Yi, Ying Shen and Liang Lin 50932

Towards Multi-modal Multi-turn Safety: From Agentic Interaction to Strategic Alignment
Han Zhu, Jiale Chen, Chengkun Cai, Shengjie Sun, Haoran Li, Yujin Zhou, Chi-Min Chan, Pengcheng Wen, Lei Li, Yike Guo and Sirui Han 50953

Case2Flow: Bridging Patient Cases and Guideline Flowcharts through Multimodal Retrieval
Jiale Wei, Yufan Chen, Alexander Jaus, Zdravko Marinov, Julian Friedrich, Simon Reiß, Jens Kleesiek and Rainer Stiefelhagen 50976

Cross-Modal Emergent Misalignment Through Narrow-Domain Fine-Tuning
Han Zhu, Chengkun Cai, Pengcheng Wen, Yanxu Zhu, Jagmohan Chauhan, Lei Li, Yike Guo and Sirui Han 50993

FG^2-GDN: Enhancing Long-Context Gated Delta Networks with Doubly Fine-Grained Control
Pingwei Sun, Yuxuan Hu, Jianchao Tan, Xue Wang, Jiaqi Zhang, Yifan Lu, Yerui Sun, Yuchen Xie and Xunliang Cai 51013

Video2Code: Generating Interactive Webpages from UI Videos via Action-Aware Revisit
Mingde Xu, Zhen Yang, Yan Wang, Yu Wang, Xijun Liu, Zijun Dou, Wenyi Hong, Xiaotao Gu, Bin Xu and Jie Tang 51024

ParentingLLM : Pacing Reasoning Maturity for Bidirectional Test-Time Scaling under Token Budgets
Seunghwan Jeong, Sugi Lee, WeiJin Park, Ikjun Yeom and Honguk Woo 51058

Evaluating and Mitigating Anti-LGBTQ Biases in German and Multilingual Language Models
Melina Morch and Daniel Braun 51072

RAG-Safety-Bench: Reliable Evaluation of Retrieval-Augmented LLM Safety
Adithiyan Rajan Indira Saravanan and Kathleen C. Fraser 51088

CoEmbod: A Long-Horizon Benchmark for Multi-Agent Embodied Cooperation with Diverse Inter-Agent Dependencies
Xuan Dong, Zhe Han, Ru Yu, Tianhao Niu, Muzhen Cai, Xinda lv, Qingfu Zhu, Weinan Zhang and Wanxiang Che 51112

DentalVoice: EMR Grounded Synthetic Dental Speech Dialogues for AI Scribes
Zihan Wang, Yi Zhou, Xuliang Yu, Hao WU, Qianyi Cai, Yi Cui, Minghao Yang and Hui Xiong 51132

In LLM Reasoning, there is Irrationality on top of Value Misalignment
Kejiang Qian and Fengxiang He 51157

AURA-KV: Adaptive Subspace Rank Approximation for Efficient Long-Context LLM Inference
Jincheng Xie, Hao Shi, Zhengyi Jiang, Zhongyi Huang, Linqi Song and Hanxu Hou 51179

Fine-Tuning for Social-Media User Simulation: A Study Across Models and Evaluators
Sneheel Sarangi, Aurélien Bück-Kaeffer, Maximilian Puelma Touzel, Zachary Yang, Jean-François Godbout and Reihaneh Rabbany 51204

Enabling Confidence-Aware Control of Effective Memorization in Large Language Models
Chenjie Ni, Omid Mersa, Zhepeng Wang, Runxue Bao, Qun Li, Shangqian Gao and Yanfu Zhang 51233

FiMA: Redundancy-Aware Description Enhancement for Gloss-Free Sign Language Translation
Peidong Liu, Lianyu Hu and Wei Feng 51247

LLUMI: Improving LLM Writing Assistance for Mental Health Support with Online Community Feedback
Jenny Jiwon Kim, Maya Ajit, Sherry Gong, Soorya Ram Shimgekar, Dong Whi Yoo, Eshwar Chandrasekharan and Koustuv Saha 51260

LaPR: Latent Preference Reasoning for Generative Personalized Retrieval
Hanze Guo, Shunyu Zhang, Jiakai Tang, Jingwei Zhuo, Yi Zhang, Xuanping Li and Xiao Zhou 51279

When Does Training Data Matter? Dynamic Data Attribution via Window-level Influence Estimation
Jie Xu, Zihan Wu, Wenbo Pan, Yulong Ming and Xiaohua Jia 51298

Plan-and-Verify Video Reward Reasoning with Spatio-Temporal Scene Graph Grounding
Hyomin Kim, Junghye Kim, Joanie Hayoun Chung, Yoonjin Oh, Kyungjae Lee, Sungbin Lim and Sungwoong Kim 51317

Lossy Orthography and Linguistic Inference: Computational Approaches to Arabic Rasm Restoration
Maged S. Al-shaibani and Irfan Ahmad 51351

Layer-wise Curriculum Learning for Efficient LLM Compression
Donggeon Lee, Dooyeon Na, Seungmin Oh and Jongbin Ryu 51372

ActionRAG: Constraint-Aware Action Retrieval for Large Action Models
Inho Lee and Yunsick Sung 51389

How Instruction Hierarchy Breaks Under Pairwise Multi-Conflict Pressure
Yerin Hwang, Jiwon Moon, Changho Lee, Kwangmin Ki, Taegwan Kang, Jinsik Lee, Honglak Lee and Kyomin Jung 51406

When Paraphrases Break Binding: A Lexical Cliff in Small Heads over Frozen CLIP Text Embeddings
Xinmiao Du, Zhuoyu Jiang and Xihong Wu 51437

Structured Inference with Large Language Gibbs
Sanghyeok Choi, Henry Gouk and Esmeralda S. Whitammer 51454

Breaking Bored? SciMemeX: Towards Automatic Generation of Scientific Memes from Research Articles
Sandeep Kumar and Asif Ekbal 51481

Revisiting the Capacity Gap in Chain-of-Thought Distillation from a Practical Perspective
Tokio Kajitsuka, Ukyo Honda and Sho Takase 51505

VinciCoder: Unifying Multimodal Code Generation via Coarse-to-fine Visual Reinforcement Learning
Xuanle Zhao, Deyang Jiang, Zhixiong Zeng, Lei Chen, Haoyue Yang, Haibo Qiu, Jing Huang, Yufeng Zhong, Liming Zheng, Yilin Cao and Lin Ma 51529

DEEPRUBRIC: Evidence-Tree Rubric Supervision for Efficient Reinforcement Learning of Deep Research Agents
Minghang Zhu, Chuyang Wei, Junhao Xu, Yilin Cheng, Zhumin Chen and Jiyan He 51545

The Analyst in the Prompt: Role, Retrieval, and Memory Biases in LLM Financial Analysis
Ahmed Asaad, Amr Mohamed, Yang Zhang and Omneya Abdelsalam 51572

Grounding What Matters: Evidence-Aware Credit Assignment for Remote Sensing Reasoning
Ziqi Miao, Zhanxu Xie, Liangzhe Han, Yuan Xiong, Chen Qian, Lijun Li, Weifeng Lv and Leilei Sun 51598

ThinkLeap: Reasoning Optimization Through Confidence-Guided Cognitive Leap Extraction
Yijia Zhu, Zhenning Shi, Junhan Shi, Xun Zhang, Yong Jiang and Qing Li 51628

Voices of Freelance Professional Writers on AI: Limitations, Expectations, and Fears
Anastasia Ivanova, Natalia Fedorova and Ekaterina Artemova 51642

Mitigating Object Hallucination in MLLMs with Global-to-Local Visual Attention Correction
Xingxing Huang, Zhiguo Yang, Jiacheng Pi, Dongsheng Xu, Ruizhi Zhong, Peipei Xu and Wenjie Ruan 51671

Small yet Assistive: Spatially-Aware Post-Training for Low Vision
Rishabh Choudhary, Shreyansh Raj, Umesh Goyal, Subh Kasyap, Shrestha Kumar, Sushovan Jena, Komal Kumar, Hisham Cholakkal and Aditya Nigam 51690

ScreenCoder: Advancing Visual-to-Code Generation for Front-End Automation via Modular Multimodal Agents
Yilei Jiang, Yaozhi Zheng, Yuxuan Wan, Jiaming Han, Qunzhong Wang, Michael R. Lyu and Xiangyu Yue 51704

Rethinking the Role of Efficient Attention in Hybrid Architectures
Ziqing Qiao, Yinuo Xu, Chaojun Xiao, Zhou su, Zihan Zhou, Yingfa Chen, Xiaoyue Xu, Xu Han and Zhiyuan Liu 51727

Beetle: A Bilingual Model Suite for Modelling Second-Language Processing
Suchir Salhan, Catherine Arnett, James A. Michaelov and Paula Buttery 51754

OSynth: A Longitudinal, Context-Sensitive Evaluation of Meeting Understanding
Ian Porada, Suraj Tripathi, Chaitanya Animesh, Leda Sari, Chaitanya Narisetty, Austin R Matthews, Bohan Yu, Ramesh L Thulasiram, Gregory Sell, KK Chin and Dmitriy Genzel 51783

STIR: Self-Distilled Latent Tools for Dynamic Control of LLM Reasoning
Zhenning Shi, Yijia Zhu, Junhan Shi, Xun Zhang, Lei Wang and Congcong Miao 51814

SDARE-Bench: Evaluating Large Language Models on Conversational Stigma Detection and Response in Dyadic and Group Dialogue
Stephanie Fong, Yiwen Jiang, Zimu Wang, Hongxi Yang, Yaling Shen, Hiu Weh Naomi Chow, Heung Ying Lai, Xiangyu Zhao, Qingyang Xu, Zhongxing Xu, Jiahe Liu, Guilherme C Oliveira, Vincent Lee, Zongyuan Ge and Dominic Dwyer 51830

HS-SFT: Hybrid Sparse Supervised Fine-tuning for Offline LLM KV Cache Eviction
Yuxin Zhang, Ruobing Xie, Wenhao Li, Jun Xia, Xiawu Zheng, Samm Sun, Di Wang and Rongrong Ji 51864

PASTABench: Proactive Assessment of Sequential Trajectories for Agent Safety
Jiapeng Sun, Yujin Zhou, Han Zhu, Pengcheng Wen, Jiayi Zhou, Sirui Han and Yike Guo . 51889

NLData2Opt: Benchmarking Parameter Derivation and Prediction for Modeling Optimization Problems from Natural Language and Data
Lijia Liu, Kyohei Atarashi, Jiyi Li, Koh Takeuchi, Shunji Umetani and Hisashi Kashima . . 51908

GRRM: Group Relative Reward Modeling for Machine Translation
Sen Yang, Shanbo Cheng, Lu Xu, Jianbing Zhang and Shujian Huang 51934

The Answer Lies Within: Self-Derived Rewards Enable Explainable Relation Extraction
Xinyu Guo, Zhengliang Shi, Minglai Yang, Mahdi Rahimi and Mihai Surdeanu 51957

Steering the Compass: Aligning Dynamic Psychological Counseling Conversations with Cognitive Behavioral Therapy Strategies
Zimu Wang, Yiwen Jiang, Xiangyu Zhao, Yaling Shen, Jiahe Liu, Stephanie Fong, Maxmartwell H Cheng, Guilherme C Oliveira, Anh Nguyen, Robert Desimone, Barnaby Nelson, Dominic Dwyer and Zongyuan Ge 51994

In-Context Learning Beyond Copying: A Training-Time Analysis of Abstractive ICL
Kerem Şahin, Sheridan Feucht, Adam Belfki, Jannik Brinkmann, Aaron Mueller, David Bau and Chris Wendler 52015

DELTA: Decoupling Latent Heterogeneity in Asymmetric Low-Rank Compression
Jialin Zhan, Fangxin Liu, Junjie Wang, Ning Yang, Li Jiang and Haibing Guan 52046

Sequential Diffusion Language Models
Yangzhou Liu, Yue Cao, Hao Li, Gen Luo, Shihao Wang, Zhe Chen, Weiyun Wang, Xiaobo Liang, Biqing Qi, Lijun Wu, Changyao Tian, Yanting Zhang, Yuqiang Li, Tong Lu, Yu Qiao, Jifeng Dai and Wenhai Wang 52058

Lost in Interpretation: The Plausibility-Faithfulness Trade-off in Cross-Lingual Explanations
Somnath Banerjee, Pranav Jha, Rima Hazra and Animesh Mukherjee 52077

MemeLens: Multilingual Multitask VLMs for Memes
Ali Ezzat Shahroor, Mohamed Bayan Kmainasi, Abul Hasnat, Dimitar Iliyanov Dimitrov, Giovanni Da San Martino, Preslav Nakov and Firoj Alam 52102

Before You Poll with LLMs: A Deliberative Diagnostic Framework
Ahmed Wali and Hassaan Tayyab 52129

ContextBias: Controlled Evaluation of Bias Persistence Under Context Shift in Text-to-Image Models
Shaghayegh Kolli, Sina Emami, Moreno D'Incà, Pouyan Nejadi, Nicu Sebe, Massimiliano Mancini and Jana Diesner 52146

Twin Worlds: Equivariance-Based Abstention for Evidence-Grounded Reasoning
Vy Nguyen, Ziqi Xu, Jeffrey Chan, Estrid He, Feng Xia, Renqiang Luo, Erik Cambria and Xiuzhen Zhang 52170

Is Alignment Enough for Helpfulness, Harmlessness, and Honesty?
Gautam Siddharth Kashyap, Mark Dras and Usman Naseem 52189

DIVA: Exploiting Cross-Step Conditional Propagation for Visual Jailbreaks in Discrete Diffusion Vision-Language Models
Guorui Song, Runqing Tang, Jingye Zhang, Luyuan Zhang, Feice Huang, Cong Ray, Guocun Wang, Dake Zhong, Choo Sin Wai, Bingquan Dai, Chuming Wang, Tongxu Lin, Wanyu Guo and Haoqian Wang 52202

A Survey on Long-Term Memory Security in LLM Agents: Attacks, Defenses, and Governance Across the Memory Lifecycle
Zehao Lin, Xixuan Hao, Renyu Fu, Shaobo Cui, Kai Chen, Chunyu Li, Zhiyu li and Feiyu Xiong 52221

GRADFILTERING: Uncertainty-Aware Gradient Signal-to-Noise Data Selection for Instruction Tuning
Zhihang Yuan, Chengyu Yue, Long Huang, Litu Ou and Lei Shi 52236

Creative Generation via Multi-Agent Debate: Does Debate Suppress Diversity?
Tien Anh Nguyen, Khanh-Binh Nguyen, Van Dai Do, Svetha Venkatesh and Hung Le 52248

PonsRAG: A Pons-Inspired RAG Bridging Cognitive Islands for Coordinated Long Narrative Reasoning
Rongchen Zhao, Yu Chen, Juyuan Wang, Zhouting Mo, Jianxing Yu, Wenqing Chen and Jingping Liu 52276

Foresight-on-Graph: Vision-Guided Knowledge Graph Reasoning with Large Language Models
Yuwei Hu, Runlin Lei, Zhewei Wei, Yongchao Liu, Chuntao Hong, Jiajun Zheng and Juelu Zhang 52294

VisCopilot: Assisting Data-Driven Article Authoring with Fact-Based and Intent-Aligned Autocompletion
Tzu-Wei Yu, Che-Wei Chang, Evelyn Yi-Ting Hung, Jia-Yu Pan and Wen-Chieh Lin 52315

Variable-size Discovery and Representation of Atomic Propositions
Francesco Ortame and Roberto Navigli 52330

Staying on Task: Defending Against Prompt Injection via Task Consistency Checking
Feiyue Xu, Jiaming Mu, Fengting Li, Bin Benjamin Zhu, Hongsheng Hu, Qiankun Liu, Shuo Wang, Chaoxiang He and Chao Feng 52349

Improving Parameter Utilization by Sharing Neural Experts Across Layers in Transformers
Dian Jiao, Jiaxin Duan and Shuai Zhao 52376

FailureMem: A Failure-Aware Multimodal Framework for Autonomous Software Repair
Ruize Ma, Yilei Jiang, Shilin Zhang, Zheng Ma, Yi Feng, Zhi Wang, Xiangyu Yue, Chuanyi Li, Vincent Ng and Lewei Lu 52387

SIMPLE: A Comparative Study of Prompting and Fine-Tuning for LLM-Based Text Simplification
Zixian Meng and Dongqiang Yang 52403

Objective vs. Search: Decomposing What Makes a Good Tokeniser
Ahmetcan Yavuz, Clara Meister and Tiago Pimentel 52425

ConceptGuard: Proactive Safety in Text-and-Image-to-Video Generation through Multimodal Risk Detection
Ruize Ma, Minghong Cai, Yilei Jiang, Jiaming Han, Yi Feng, Yingshui Tan, Bo Zhang and Xiangyu Yue 52445

Think Before You Link: Rarity, Reasoning, and Retrieval in Multilingual Entity Linking
Parinthapat Pengpun, Simran Khanuja and Graham Neubig 52464

Apples on the Table? Evaluating Text-Guided 3D Scene Synthesis via Fine-Grained Constraint Verification
Minseok Kang, Dongwook Choi, Gyeom Hwangbo, Seungwon Lim, Kai Tzu-iunn Ong and Jinyoung Yeo 52487

AsymSpec: Context-Asymmetric Speculative Decoding for Agentic LLMs
Sheng Liang, Yongyue Zhang, Nathanael Brian, Hang Lv, Hao Wang, Chen Zhang and Yong Liu 52532

Guiding Graph: Explicit Query-Structure Guidance for Knowledge Graph Question Answering
Dehao Tao, Guangjie Liu, Jun Zhou, Junhao Chen, Wei Zheng, Rui Chen, Chen Junren, Yongfeng Huang and Minghu Jiang 52549

The Benchmark Illusion: Pruned LLMs Can Pass Multiple Choice but Fail to Answer
Rui Wen, Lu Sun, Jiayang Liu, Zesheng Xu, Tianshuo Cong and Zheng Li 52565

A Benchmark for Arabic Lemma-Level Readability Assessment
Khalid N. Elmadani, Mostafa Saeed, Hanada Taha and Nizar Habash 52576

Imagine Before You Predict: Interleaved Latent Visual Reasoning for Video Event Prediction
Tianxiang Jiang, Linquan Wu, Sheng Xia, Songze Li, Ziang Yan, Haoyu Yang, Yu Qiao and Yi Wang 52600

How Do Language Models Represent and Use Phonological Information for Allomorph Selection?
Sangwoo Kim and Sangah Lee 52622

PR-LoRA: Direction-Regularized Low-Rank Adaptation for Behavioral Alignment in Large Language Models
Yikai Zhang, Rui Xu, Aili Chen, Haoyu Zhang, Libo Wu, Xu Yinghui and Yanghua Xiao . 52640

CebIDN: A Computational Framework for Cebuano Idiom Identification, Disambiguation, and Normalization
Adrian P. Vaflor and Robert R. Roxas 52655

Knowing Isn't Always Saying: When Do Spatial Encodings Reach Answers in Vision-Language Models?
Zeyu Wang and Xinming Xu 52683

DeARAG: Decoupled Agentic RAG for On-Device and Precise Multi-Hop Question Answering
Qiuchang Shi, Tianqi Liu, Feng Zhang, Peng Chen, Jianjun Liu and Gaoyun An 52706

DiaLLM: An Investigation into the Robustness-Generation Gap in English Dialect Adaptation
Jordan Painter, Dipankar Srirag, Adarsh Kappiyath, Diptesh Kanojia, Aditya Joshi and Lu Yin 52728

Anomaly Detection using Source-Reversed Flow Matching
Amine Youcef Khodja, Jeremie Pantin, Gaël Dias and Fabrice Maurel 52746

Label-Free Post-hoc Alignment of Frozen Language Model Hidden States for Sentence-Similarity Retrieval
GangHo Lee and Yong-Gu Lee 52768

Automatic Configuration of LLM Post-Training Pipelines
Channe Chwa, Xinle Wu and Yao Lu 52787

What Does a Language Model Know If It Has Never Seen the Future? Temporal Isolation as a Natural Experiment in Knowledge Attribution
Azizur Rahman, Rushikesh Hiray, Shubham Srivastava, Mohammad Hijjawi and Laraib Hasan 52814

Harness Updating Is Not Harness Benefit: Disentangling Evolution Capabilities in Self-Evolving LLM Agents
Minhua Lin, Juncheng Wu, Zijun Wang, Zhan Shi, Yisi Sang, Bing He, Zewen Liu, Tianxin Wei, Zongyu Wu, Zhiwei Zhang, Dakuo Wang, Xiang Zhang, Benoit Dumoulin, Cihang Xie, Yuyin Zhou, Suhang Wang and Hanqing Lu 52829

Signals from the Future: Future-aware Semi-Autoregressive Decoding for Diffusion Language Models
Yangryeol Park, Jeonhwi Kang, Kunhui Lee, Hyeonju Kim, Jinyoung Choi and Seung-Hoon Na 52853

TimelineAgent: An LLM-based Framework for Query-Centric Event Timeline Extraction Across Multiple Documents
Qingkai Min, Yunxin Bao, Yifeng Lu, Zhenyu Feng, Yulong Chen, Andreas Vlachos and Yue Zhang 52872

DEAM: Diverse Embeddings for Arabic Multimodal Retrieval
Omer Nacar, Fatimah Mohamed Emad Elden, Rahaf Saeed Al Hallay, Mona A.Azim, Aya Hamod, Mohamed Zaytoon, Nour Qandos, Mohamed Abdelaziz, Israa Elhosiny, Yousra Alaa El-Ghawi, Serry Sibaee, Hadi Hamoud, Norah Naji Al-Zaid, Muhammed Ragab and Khloud Al Jallad 52897

Exploring Autonomous Agentic Data Engineering for Model Specialization
Yujie Luo, Xiangyuan Ru, Jingsheng Zheng, Jingjing Wang, Yuqi Zhu, Jintian Zhang, Runnan Fang, Kewei Xu, Ye Liu, Zheng Wei, Brucebian, Zang Li and Shumin Deng 52914

Multi-Faceted Interactivity Alignment in Full-Duplex Speech Models
Atsumoto Ohashi, Neil Zeghidour, Alexandre Défossez and Eugene Kharitonov 52949

Are We Measuring Strategy or Phrasing? The Gap Between Surface- and Approach-Level Diversity in LLM Math Reasoning
Sangmook Lee, Minbeom Kim, Jeonghye Kim, Dohyung Kim, Sojeong Rhee and Kyomin Jung 52966

Categorize Early, Integrate Late: Divergent Processing Strategies in Automatic Speech Recognition
Nathan Roll, Pranav Bhalerao, Martijn Bartelds, Arjun Pawar, Yuka Tatsumi, Tolúlopé Ògúnrèmí, Chen Shani, Calbert Graham, Meghan Sumner and Dan Jurafsky 52997

GUARD: Natural Forgetting in Large Reasoning Models via Guided Answer-Reasoning Distillation
Zeyu Yan, Guanghao Zhou, Minghui Qiu, Ming Gao and Cen Chen 53013

SHADOWBENCH: Toward Reliable Automatic Evaluation of Semantic Alignment in Autoformalization
Hojae Han, Jongyoon Kim, Sanghyeok Park, Dongwook Cheon, Yeachan Park, Myeong Jae Jeon, SunJong Choe, Soonho Kong, Wonseok Hur, Seung-Won Hwang and Donghoon Hyeon 53031

MILP-Solve-Free Selection of LLM-Generated MILP Programs via Multi-Signal Structural-Semantic Consistency
Shuli Zeng, Guozhu Liu, Yinghao Chen, Wantong Xie, Ji Liu, Shaojie Tang, Feng Wu, Sijia Zhang and Xiang-Yang Li 53068

Another Turn, Better Output? A Turn-Wise Analysis of Iterative LLM Prompting
Shashidhar Reddy Javaji, Bhavul Gauri and Zining Zhu 53085

The Company You Keep: How LLMs Respond to Dark Triad Traits
Angelica Henestrosa, Zeyi Lu, Pavel Chizhov and Ivan P. Yamshchikov 53111

MemMA: Coordinating the Memory Cycle through Multi-Agent Reasoning and In-Situ Self-Evolution
Minhua Lin, Zhiwei Zhang, Hanqing Lu, Hui Liu, Xianfeng Tang, Qi He, Xiang Zhang and Suhang Wang 53127

Dynamic Rollout Editing for Reducing Overthinking in RL-Trained Reasoning Models
Zihao Wei, Wenjie Shi, Liang Pang, Jingcheng Deng, Shicheng Xu, Shasha Guo, Zenghao Duan, Jiahao Liu, Jingang Wang, Huawei Shen and Xueqi Cheng 53151

Chronos: Chronology-Grounded Continued Pre-training for Language Model Adaptation to Materials Science
Sangyun Kim, Junho Kim and SangKeun Lee 53173

To Answer or Not to Answer? Reliability-Oriented Reinforcement Learning for Visual Language Models
Paul Jian, Julian Wu, Leo Wei, Dylan Chen, Garvin Guo, Qingpei Guo and Jianing Li 53194

A Robust Evaluation of Probe Robustness: Lessons for Reliable OOD Uncertainty Quantification
Joe Stacey, Hadas Orgad, Kentaro Inui, Benjamin Heinzerling and Nafise Sadat Moosavi . 53221

Latent Mechanisms of Language Control in Multilingual Language Models
Ryo Mitsuhashi, Sabri Boughorbel and Majd Hawasly 53250

Latent-GRPO: Group Relative Policy Optimization for Latent Reasoning
Jingcheng Deng, Zihao Wei, Liang Pang, Junhong Wu, Shicheng Xu, Zenghao Duan and Huawei Shen 53273

Latent-SFT: LLM Latent Reasoning as Chain of Superposition
Jingcheng Deng, Liang Pang, Zihao Wei, Shicheng Xu, Zenghao Duan, Kun Xu, Yang Song, Huawei Shen and Xueqi Cheng 53289

Genesis: Visual Augmentation Primitives for Multimodal Chain-of-Thought Reasoning
Fan Yang, Lifeng Chen, Yipu Wang, Han Wang, Feilong Chen, Zhiyang Chen, Yidong Ma, Chaoyang Zhao, Hanqi Jiang, Yibo Chen, Xu Tang, Yao Hu, Ming Tang, Xiaojie Jin, Xiaolong Zheng and Jinqiao Wang 53310

KV-Retriever: Aligning KV Cache Eviction with Downstream Task Loss
Wenhao Gao, Binqin Shi, Wei Jia, Aiwei Liu, Linhao Zhang, Caixia Yuan, Xiaojie Wang and Zhou Xiao 53329

Model Interruption Bench: Evaluating Real-Time Error Correction in Full-Duplex Voice Agents
Shih-Yun Shan Kuan, Cheng-Han Chiang, Guan-Ting Lin and Hung-yi Lee 53341

What Transfers from Text to Vision? Capability Scaling Laws and Transfer Dynamics for VLMs
Qiang Wang, Ziran Li, Zhengyu Chen, Shanglin Lei, Borun Chen, Jingang Wang and Xunliang Cai 53357

INTENT-AS-A-TOOL Makes it Easy to Track Agentic Misalignment
Yutong Zhang, Jianshuo Dong, Peng Xu, Long Wang, Jie Zhang, Tianwei Zhang, Xiaoping Zhang and Han Qiu 53384

Beyond Superficial Unlearning: Sharpness-Aware Robust Erasure of Hallucinations in Multimodal LLMs
Xianya Fang, Feiyang Ren, Xiang Chen, Yu Tian, Zhen Bi, Haiyang Yu and Sheng-Jun Huang 53419

Steering Vector Fields for Context-Aware Inference-Time Control in Large Language Models
Jiaqian Li, Yanshu Li and Kuan-Hao Huang 53438

TETHER: Training-Free, Elastic, ToC-Anchored Hierarchy Extraction and Recovery in Design-Driven Long Documents
Yu Zou, Yan Chen, Jialei Zeng, Xiaorui Zhou, Aixi Zhong, Qi Zhou, Haoran You, Ke Jiang and Changyu Hu 53469

HOMURA: Taming the Sand-Glass for Time-Constrained LLM Translation via Reinforcement Learning
Ziang Cui, Mengran Yu, Chenyu Shi, Yingxuan Shi and Tianjiao Li 53480

Discourse-Aware Policy Analysis with Argumentation: A Hybrid LLM–Symbolic Framework for Disaster Governance
Stylianos Loukas Vasileiou and Olga Derendiaeva 53504

KnowGPO: Knowledge-Grouped Policy Optimization for Search Agents
Siqi Chen, Zhengxi Lu, Yuchen Yan, Xingyu Wu, Qipeng Chen, Xin Zhang, Aiting Liu, Chao Deng, Jie Liu, Jin Ma, Weiming Lu, Jun Xiao and Yongliang Shen 53526

SMART: Sink-based Modality-Aware Redistribution of Transformer Attention
Dian Yu, Yao Yao and Hai Zhao 53547

KidnapRAG: A Black-Box Attack for Hijacking Reasoning in Agentic Retrieval-Augmented Generation Systems
Chanwoo Choi, Euntae Kim, Kyuho Lee, Youngsam Chun, Jinhee Jeong, Eunmi Kim, Myunggyo Oh, Junseo Jang and Buru Chang 53561

RAPA: Recursively Aligned Pathway Adaptation of Large Language Models
Sol Namkung, Sunghyeon Woo, Inho Jeong and Dongsuk Jeon 53591

MemeArgs: Vision-Language Argument Mining in Persuasive Memes
Khoi P. N. Nguyen, Trung N. Nguyen, Terrence Li, Teerth Patel and Vincent Ng 53619

Concepts Whisper: Spectral Anti-Concentration and the Dual Geometry of Transformer Representations
Pratyush Acharya, Nuraj Rimal and Habish Dhakal 53643

Seeing with You: Perception-Reasoning Co-evolution for Multimodal Reasoning
Ziqi Miao, Haonan Jia, Lijun Li, Chen Qian, Yuan Xiong, Wenting Yan and Jing Shao 53668

Norms of Error Terms in Transformer Language Models Predict Human Reading Time Beyond Surprisal
Ziyuan Cao and William Schuler 53691

Beyond the Efficiency Paradox: Unlocking Multimodal Subspaces via Cumulative Information Span
Xiuze Zhou, Xiaowen Chu and Xuming Hu 53700

GreenSwap: Verb Choice or Output Length? A Controlled Decomposition of Prompt-Level Energy Savings in LLM Inference
Ahsan Farabi, Israt Khandaker, Md Abdul Ahad Minhaz, Zannatul Zahan Meem, Ibrahim Khalil Shanto and Shahriar Mahmud 53711

Pruning Laws for Large Language Models
Ayan Sengupta, Siddhant Chaudhary and Tanmoy Chakraborty 53725

Few-shot Semantic Recovery Attacks on Image Embeddings
Yiyi Chen, Qiongkai Xu and Johannes Bjerva 53745

CRICO: A Certified Robustness Framework via Input-Dependent Covariance Optimization
Shudi Guo, Bin Liu, Shihai Wang, Tengfei Shi, Yu Liu and Haoran Li 53768

Beyond Consensus: Perspectivist Modeling and Evaluation of Annotator Disagreement in NLP
Yinuo Xu and David Jurgens 53785

A Gradient Perspective on RLVR Stability and Winner-Advantage Policy Optimization
Prasanth Yss, Zhichen Ren, Rasa Hosseinzadeh, Ilan Gofman, Yuqi Chen, Zhaoyan Liu, Guangwei Yu, Jesse C. Cresswell and Satya Krishna Gorti 53806

Dreaming Up Authors: Factual Prevalence and LLM Hallucinations in Academic Authorship Attribution
Sijia Zhong, Tamir Yirga, Yijia Ma, Thai Le and Yifan Hu 53827

When Facts Change: Probing LLMs on Evolving Knowledge with evolveQA
Nishanth Sridhar Nakshatri, Shamik Roy, Manoj Ghuhan Arivazhagan, Hanhan Zhou, Vinayshekhar Bannihatti Kumar and Rashmi Gangadharaiah 53864

OVO-S-Bench: A Hierarchical Benchmark for Streaming Spatial Intelligence in Multimodal LLMs
Yifei Li, Pengyiang Liu, Yuhang Zang, Zhongyue Shi, Qi Fu, Hongye Hao and Jiwen Lu . 53894

Compositional SVG Generation via VLM-Driven Hierarchical Semantic Parsing
Sehwan Park, Taehoon Kim, Geonhee Han, Dohyun Kim, Seung Wook Kim and Paul Hongsuck Seo 53942

The Trigger Is in the Embedding: Supervision-Free Discovery and Diagnosis of Universal Adversarial Triggers in Fact Verification
Michele Luca Contalbo, Matteo Paganelli, Paolo Papotti and Francesco Guerra 53968

HPP: Hierarchical Programmatic Probing for Long Video Understanding by Decoupling Perception and Reasoning
Awais Rauf, Ahmed Hasssan and Gregory Slabaugh 53993

DiffNeCTIS: Diffusion-based Nested Compound Type Identification for Sanskrit
Pretam Ray, Sinchani Chakraborty, Sriram Krishnan, Manoj Balaji Jagadeeshan and Pawan Goyal 54021

SeMLI-RAG: Semantics-Guided Multimodal Late Interaction for Retrieval-Augmented Radiology Report Generation
Jiawei He, Weibo Feng, Sohrab Ferdowsi, Anthony Yazdani, Boya Zhang, Osman Alperen Çinar-Koraş, Sameh Khattab, Jean-Philippe Corbeil, Henning Müller and Douglas Teodoro 54040

An Embarrassingly Simple Data Selection Strategy for Million-Scale Multi-Modal Instruction-Tuning
Mengyao Lyu, Yan Li, Huasong Zhong, Hui Chen, Jungong Han, Guiguang Ding and Zhenheng Yang 54056

Evolutionary Soups: Evolving Mixture-of-Experts for Multi-Objective LLM Alignment
Lingxiao Kong, Steffen Staab, Cong Yang, Oya Beyan and Zeyd Boukhers 54100

Joint Constraint Aggregation for Qualifier-Sensitive Retrieval-Augmented QA
Jennifer Hsia, Zachary Chase Lipton and Aarti Singh 54126

AGE-Sampler: Adaptive-granularity Graph-enhanced Event Sampling for Long-Video Understanding
Yanzi Wang, Tao Dai, Yaohua Zha and Shu-Tao Xia 54141

RoboReason: Boosting Robotics-Relevant VLM Reasoning Through Scalable Data Generation
Sagnik Anupam, Utkarsh Lal, Aneesh Edara, Botong Zhang, Yao Tang, Jiatao Gu, Dinesh Jayaraman and Osbert Bastani 54156

AnchorLens: A Shared Coordinate System for Interpreting Intermediate Representations in Large Language Models
Giuseppe Trimigno, Gianfranco Lombardo and Stefano Cagnoni 54183

Temporal Leakage in Financial News NLP: A Multi-Architecture Audit with a Regime-Specific M&A Signal
Chenhao Xue, Raslen Guesmi, Siwei Feng, Yucheng Gong, Jacob Xavier Sundram, Jordan Pang, Lan Wang and Julian Kaljuvee 54209

AtomMem: Building Simple and Effective Memory System for LLM Agents via Atomic Facts
Yanyu Yao, Shangze Li, Zhi Zheng, Hui Zheng, Qi Liu, Tong Xu and Enhong Chen 54243

Clinically Grounded Privacy Evaluation of Medical LMs
Sasha Ronaghi, Sana Tonekaboni, Lena Stempfle, Vivian Utti, Jordan Li Cahoon, Nathaniel Hendrix, Ayin Vala, Marzyeh Ghassemi and Emily Alsentzer 54266

Watch Your Norms: On Test-Time Expert Reduction in Multilingual MoEs
Benedikt Ebing and Goran Glavaš 54296

CauTab: Counterfactual Tabular Data Synthesis with Functional and Causal-Dependency Constraints
Xingyu Yang, Mingchuan Ma, Dexuan Xu, Xiangyang Long and Yu Huang 54342

An Interpretable Linguistically-Grounded Analysis of Scene Saliency in Movie Screenplays
Uday Bindal, Ishaan Karan, Manan Garg, Ekansh Goyal, Rahul Mishra and Rajakrishnan P Rajkumar 54364

Near-Miss Context Contamination and Language-Asymmetric Evidence Arbitration in LLMs
Jebish Purbey, Sanjeeb Prasad Panday, Barnnita Shrestha, Bikash Balami, Aakash Pandey and Bibek Khanal 54388

PROVENA-MED: Benchmarking Provenance Faithfulness in Multimodal Clinical Reasoning
J. Ben Tamo, Yining Yuan, Benoit Louis Marteau, Micky C. Nnamdi, Yuxing Lu and May Dongmei Wang 54405

Enhancing linear attention with residual learning
Xunhao Lai, Jialiang Kang, Jianqiao Lu, Pengyu Zhao and Tong Lin 54422

AdaPath: Query-Adaptive Path-Finding via Path-Bank for Multi-Hop Implicit Biomedical KGQA
Jun Hyeong Kim, Dongki Kim, Yinhua Piao and Sung Ju Hwang 54435

Context Compression for LLM Agents: A Survey of Methods, Failure Modes, and Evaluation
Yifei Wang, Ziteng Wang, Yuling Shi, Silin Chen, Xinrui Wang, Yueqi Wang, Beijun Shen, Linjing Li, Xiaodong Gu, Julian McAuley and Daniel Dajun Zeng 54460

Estimating Surprisal from the Speech Signal
Zachary Hopton, Samuel Kiegeland, Francesco Ignazio Re, Patrick Haller, Matthew B. Winn, Ryan Cotterell and Eleanor Chodroff 54489

Entropy Equilibrium Sampling: Tuning-Free Truncation for Language Model Decoding
Xiaodong Cai, Hai Lin, Shaoxiong Zhan, Weiqi Luo, Hai-Tao Zheng, Hong-Gee Kim, XiuTeng Zhou, Chunxia Ma and Yu Yang 54510

Revisiting the Estimation of Unigram Surprisal in Computational Psycholinguistics
Zachary Hopton, Francesco Ignazio Re, Samuel Kiegeland, Andreas Opedal, Eleanor Chodroff and Ryan Cotterell 54526

AgentCollab: A Self-Evaluation-Driven Collaboration Paradigm for Efficient LLM Agents
Wenbo Gao, Renxi Liu, Xian Wang, Guo Fang, Shuai Yang, Chen xi, Hui-Ling Zhen, Hanting Chen, Weizhe Lin, Xiaosong Li and Yaoyuan Wang 54555

MobileWorldBench: Towards Semantic World Modeling For Mobile Agents
Shufan Li, Konstantinos Kallidromitis, Akash Gokul, Yusuke Kato, Kazuki Kozuka and Aditya Grover 54571

ACTOR: Agentic CT Reasoning with Radiologist-Aligned Actions
Xingyu Qian, Yu Sun, Songyou Li, Zongzhao Li, Jiacheng Cen, Wenqian Wu, Tingyang Xu, Yu Rong, Deli Zhao and Wenbing Huang 54593

VISTA: Structure-Aware Training-Free Editable Diagram Generation from Natural Language
Dongqi Han, Bojie Lin, Junhan Wang and Jiefu Luo 54612

DynaWeb: Model-Based Reinforcement Learning of Web Agents
Hang Ding, Peidong Liu, Junqiao Wang, Ziwei Ji, Meng Cao, Rongzhao Zhang, Lynn Ai, Eric Yang, Tianyu Shi and Lei Yu 54628

SAGE: Semantic Attribute Graphs for Multi-Entity Visual Retrieval
Yongjoo Kim, Mincheol Kwon, Seonga Choi, Minseung Lee, Kyeong-Jin Oh, Hyunyoung Lee, Yunsu Choi and Jungbeom Lee 54643

dwDPO: Training Multi-Turn LLM Agents via Divergence-Weighted Direct Preference Optimization
Jinghao Lin, Yuhang Wu, Yuanqing Wang, Yuchen Li, Tianxingjian Kang, Baris Coskunuzer and Xiawu Zheng 54657

Leveraging Routing Dynamics in Mixture-of-Experts Models for Efficient Language Adaptation
Aditi Khandelwal, Marius Mosbach, Verna Dankers, Siva Reddy and Golnoosh Farnadi . . . 54675

Multi-Facility Location Mechanism Design with Large Language Models
Nguyen T Thach, Fei Liu, Houyu Zhou and Hau Chan 54691

Qwen-Image-Layered 1.5: Unified Image Generation and Editing with Native Transparency Awareness
Shengming Yin, Zekai Zhang, Yixian Xu, Kaiyuan Gao, Zhendong Wang, Lihan Jiang, Ningyuan Tang, Tianhe Wu, Xiaoyue Chen, Yanran Zhang, Zihao Liu, Zikai Zhou, Jiahao Li, Xiao Xu, Kun Yan, Yuxiang Chen, Yilei Chen, Yan Shu, Jie Zhang, Heung-Yeung Shum, Lionel Ni, Jingren Zhou and Chenfei Wu 54711

Continuous-Utility Direct Preference Optimization
Muhammad Ahmed Mohsin, Muhammad Umer and Emily Fox 54724

Improving LLMs via Validator-to-Generator Alignment
Juan Diego Rodriguez, Jocelyn Zhang, Katrin Erk and Greg Durrett 54769

On the Design Fundamentals of Pixel Text Representation Learning
Chaohao Yuan, Ruifeng Yuan, Zhuoxu Huang, Yu Rong, Hong Cheng, Hou Pong Chan and Chenghao Xiao 54787

EpiPersona: Persona Projection and Episode Coupling for Pluralistic Preference Modeling
Yujie Zhang, Weikang Yuan, Zhuoren Jiang and Pengwei Yan 54804

Q-Strata: Hierarchical Bit Allocation for Mixed-Precision Quantization of Mixture-of-Experts LLMs
Deokjae Lee, Sihun Chu and Hyun Oh Song 54824

AtlasNLP: A Country-Aware Atlas of Dataset Representation in NLP
Joan Nwatu, Tsedeniya Solomon Amare, Longju Bai, Bontu Fufa Balcha, Zayd Bashir, Angana Borah, Zara Burzo, Yubin Choi, Naihao Deng, Samika Gupta, Michel Faloughi, Claude Kwizera, Ziqiao Ma, Cynthia Yacel Fuertes Panizo, Ellie Seehorn, Hui Shen, Jiayi Tang, Zesen Zhao, Boyuan Zheng and Rada Mihalcea 54839

Nemotron-Labs-Diffusion: A Tri-Mode Language Model Unifying Autoregressive, Diffusion, and Self-Speculation Decoding
Yonggan Fu, Lexington Whalen, Abhinav Garg, Chengyue Wu, Maksim Khadkevich, Nicolai Oswald, Enze Xie, Daniel Egert, Sharath Turuvekere Sreenivas, Shizhe Diao, Chenhan D. Yu, Ye Yu, Weijia Chen, Sajad Norouzi, Jingyu Liu, Shiyi Lan, Ligeng Zhu, Jin Wang, Jindong Jiang, Morteza Mardani, Mehran Maghoumi, Song Han, Ante Jukić, Nima Tajbakhsh, Jan Kautz and Pavlo Molchanov 54868

HG-Bench: A Benchmark for Multi-Page Handwritten Answer-Region Grounding in Automated Homework Assessment
Chuangxin Zhao, Boyan Shi, Yanling Wang, Yijian LU, Canran Xiao, Jiali Chen, Jun Xia, Yan Wang, Ji Qi and Juanzi Li 54888

Fine PT-PT Web: A High-Quality 41 Billion Tokens Data Collection of the European Portuguese Web
Gonçalo Vinagre, Rui Pedro Guerra, Pedro Gomes, Miguel Moura Ramos, Duarte Miguel Alves, Afonso Simplício, Diogo Tavares, David Semedo, Daniel Gomes and Joao Magalhaes 54905

MMEmb-R1: Reasoning-Enhanced Multimodal Embedding with Pair-Aware Selection and Adaptive Control
Yuchi Wang, Dingkang Yang, Haiyang Yu, Weikang Bian, Jiefeng Long, Liang Xiao, Chao Feng and Hongsheng Li 54921

When Personas Become Harmful: Guiding LLM with Global Citizenship Values Alignment
Siqi Shen, Lajanugen Logeswaran, Moontae Lee, Honglak Lee and Rada Mihalcea 54944

Beyond Epistemic Collapse: Disagreement-Aware Scientific Retrieval-Augmented Generation
Animesh Mishra, Krishang Sharma and Sonia Khetarpaul 54964

Oblivion: Self-Adaptive Agentic Memory Control through Decay-Driven Activation
Ashish Rana, Chia-Chien Hung, Qumeng Sun, Julian Martin Kunkel and Carolin Lawrence54986

ReproRepo: Scaling Reproducibility Audits with GitHub Repository Issues
Shanda Li, Qiuhong Anna Wei, Jingwu Tang, Valerie Chen, Nihar B Shah, Tim Dettmers, Yiming Yang and Ameet Talwalkar 55018

Why Thinking Hurts: Diagnosing and Rectifying Linguistic Inertia in Large Language Models for Recommendation
Luankang Zhang, Yonghao Huang, Hang Lv, Xuyang Zhi, Mingjia Yin, Yuyang Ye, Wei Guo, Hao Wang and Enhong Chen 55037

DynaTokens: Controlling Token Dynamics for Continual Video-Language Understanding
Toan Nguyen, Yang Liu, Celso M de Melo and Flora D. Salim 55050

Persuasion Index: A Theory-Guided Framework for Persuasion Analysis
Liancheng Gong, Zhiyang Wang, Yiwei Xu and Julia Mendelsohn 55071

Text4Speech: Text-Driven Training for Massively Multilingual Speech Understanding
Hannah Sterz, Anna Korhonen, Ivan Vulić and Goran Glavaš 55111

AnchorKV: Training-Free KV Cache Compression via Joint Coverage Diversity and Semantic Importance Scoring
Mengbo Wang, Zhuochen Fan, Dayu Wang, Zirui Liu, Peiyuan Zong, Guorui Xie, Zeyu Luan, Qing Li and Yong Jiang 55127

UDSS-BWE: Uncertainty- and Decision-Science Inspired Swin BandWidth Extension
Tarikul Islam Tamiti, Sajid Fardin Dipto, David Vergano, Luke Baja-Ricketts and Anomadarshi Barua 55142

What AI Gets to See: Auditing the Loss of Multilingual Community Knowledge in Machine-Facing Metadata on Wikimedia Commons
Yihan Yu and Aylin Caliskan 55176

PRISM: Prompt Representation Internalization via Selective Compression into LoRA Weights
Eshita Jha, Mingxian Wang, Andrei Barbu and Chetan Aggarwal 55189

Breaking Causality in LLM Training with Sparsity and Quantization
George Retsinas, Anjulie Agrusa, Oleg Rybakov, Nicholas Knight, Mikail Khona, Chong Yu, Tijmen Blankevoort and Jeff Pool 55202

Narrative Flattening: How Post-Training Compresses Thematic, Affective, and Stylistic Variation in LLM Fiction
Zehan Li, Yutong Zhu, Siyang Wu, Honglin Bao and James Evans 55222

How Cognitive Biases Propagate in LLM Clinical Differential Diagnosis
Soma Bálint Nagy and Richard Farkas 55252

CERVINO: A Large-Scale Eye-Tracking Resource of Italian Legal Texts
Calogero Jerik Scozzaro, Matteo Delsanto, Sara Pasquet, Luisa Revelli, Antonio Mastropaolo and Daniele P. Radicioni 55269

CIVI: A Framework for Diagnosing Search Agent Failures in Civic Information
Dingying Liu, Yunshun Zhong, Wentao Zhang and Yiyuan Li 55284

How Do Large Language Models Learn Concepts During Continual Pre-Training?
Barry Menglong Yao, Sha Li, Yunzhi Yao, Minqian Liu, Zaishuo Xia, Qifan Wang and Lifu Huang 55302

Discover, Then Diversify: Phase-Aware Control of Tree-of-Thought Search
Atkia Mahila, Avinash Maurya, M. Mustafa Rafique and Bogdan Nicolae 55326

Do Language Models Follow Occam's Razor? An Evaluation of Parsimony in Inductive and Abductive Reasoning
Yunxin Sun and Abulhair Saparov 55345

Agentic Reinforcement Learning for Search Misaligns Instruction-Tuning
Yushi Yang, Shreyansh Padarha, Sarah Ball, Andrew Lee and Adam Mahdi 55367

Do Language Models Reason Across Languages?
Yan Meng, Wafaa Mohammed and Christof Monz 55396

Prototype-Based Dynamic Steering for Large Language Models
Ceyhun Efe Kayan and Li Zhang 55412

Evaluating the Utility of Grounding Documents with Reference-Free LLM-based Metrics
Yilun Hua, Giuseppe Castellucci, Peter Schulam, Heba Elfardy and Kevin Small 55426

Beyond Surface Forms: A Comprehensive, Mechanism-Oriented Taxonomy of Indirect Linguistic Encoding for LLM-Based Coded Language Detection
Hamid Reza Firoozfar, Mohammadsadegh Abolhasani, Reza Mousavi and Paul Jen-Hwa Hu55450

Probing Factual Knowledge Transfer with Training Data Interventions
Romina Oji, Marc Braun, Marcel Bollmann, Marco Kuhlmann and Jenny Kunz 55474

Probing Ambiguity: Adversarial Intent Resolution for Robust Text-to-SQL
Youbing Yin, Xujun Peng, Muhammad Sharif Uddin, Erin Babinsky, Daben Liu and Anoop Kumar 55493

What Are LLM Agents Actually Learning on Low-Budget Molecular Optimization Benchmarks?
Jiayu Qin, Zhengquan Luo, Changyou Chen and Zhiqiang xu 55505

ExecRetrieval: Measuring the Functional-Correctness Gap in Code-Embedding Retrieval
Aaryan Kapoor and Md Abdullah Al Hafiz Khan 55517

SketchVLM: Vision language models can annotate images to explain thoughts and guide users
Brandon Collins, Logan Bolton, Hung Nguyen, Mohammad Reza Taesiri, Trung Bui and Anh Totti Nguyen 55534

Scaling Near-Optimal SFT-RL Annotation Budget Allocation from Small to Large LLMs
Jingtan Wang, Arun Verma, Xiaoqiang Lin, Zhengyuan Liu, Nancy F. Chen, Daniela Rus and Bryan Kian Hsiang Low 55601

MacAgentBench: Benchmarking AI Agents on Real-World macOS Desktop
Yikun Fu, Bowen Fu, Zhenyu Wu, Shuang Cheng, Xiaowei Sun, Bowen Yang, Zehao Li, Yibo Zhao, Zichen Ding, Zhoumianze Liu, Shijie Wang, Biqing Qi and Bowen Zhou 55631

When Tokenization is Secretly Output Supervision
Tanja Baeumel, Josef Van Genabith and Simon Ostermann 55664

Delta-Crosscoder: Robust Crosscoder Model Diffing in Narrow Fine-Tuning Regimes
Aly M. Kassem, Thomas Jiralerspong, Negar Rostamzadeh and Golnoosh Farnadi 55698

Evaluating LLMs on Conversational Text-to-SQL under Chain Ambiguity and Intent Drift
Yujia Liu, Jiayan Lin, Zijin Hong, Zheng Yuan, Shengyuan Chen, Hao Chen, Qinggang Zhang, Xiao Huang and Feiran Huang 55720

Towards Understanding the Robustness of Self-Consuming Generative Models
Aobo Chen, Wei Qian, Chenglin Miao and Mengdi Huai 55740

mPhonBench: Evaluating LLMs On Multilingual Phonology
Lennart Keller and Goran Glavaš 55752

Beyond Sequence Order: Syntax-Informed Positional Embeddings for Transformers
Haris Riaz, Hyungji Kim and Mihai Surdeanu 55784

ParsVoice: A Large-Scale Multi-Speaker Persian Speech Corpus for Text-to-Speech Synthesis
Mohammad Javad Ranjbar Kalahroodi, Azadeh Shakery and Heshaam Faili 55804

How broad is that claim? Mapping Generalisation in NLP Research
Chenxin Diao, Nataliya Stepanova and Emily Allaway 55820

How Do Prompt Variations Affect Energy Consumption in On-Device LLMs?
Wei Hu, Xiaolong Tu, Dawei Chen, Yitao Chen, Kyungtae Han and Haoxin Wang 55852

Safety Training Modulates Harmful Misalignment Under On-Policy RL, But Direction Depends on Environment Design
Leon Eshuijs, Shihan Wang and Antske Fokkens 55879

API Benchmark Scores Do Not Reliably Transfer to Chatbot Interfaces
Jennifer Wang, Joachim Baumann, Daniel E. Ho and Sanmi Koyejo 55909

Information-Gain Rewards over Diversity-Pruned Tests: GT-Anchored Verifier Co-Training for Reliable Code Generation
Ana Nunez and Peyman Najafirad 55943

Disentangling Statistical Preemption from Entrenchment in Language Models' Avoidance of Overgeneralization
Yixuan Wang, Freda Shi and Kanishka Misra 55959

Evaluating LLM Autograders Without Ground Truth
Huijun Mao and Christopher J Piech 55976

JURY-RL: Votes Propose, Proofs Dispose for Label-Free RLVR
Xinjie Chen, Biao Fu, Wu Jing, Guoxin Chen, Xinggao Liu, Dayiheng Liu and Minpeng Liao 55987

Jury Duty: Calibration and Orientation Failures in MLLM-as-a-Judge Under Cultural Ambiguity
Daniel Lee, Harshit Sharma, Eunkyu Park, Pranav Narayanan Venkit, Jeonghwan Kim, Kah Mun Chia, Andreas Vlachos and Shafiq Joty 56015

Do LVLMs Know When They Are Wrong? Trajectory-Based Correctness Estimation from a Single Forward Pass
Awal Ahmed Fime, Md Zarif Hossain and Ahmed Imteaj 56033

LLMpedia: A Transparent Framework to Materialize an LLM's Encyclopedic Knowledge at Scale
Muhammed Saeed and Simon Razniewski 56051

The Interlingua Hypothesis: LLMs Translate via a Latent Task-agnostic Feature Space
Jacob Brinton, Jannik Brinkmann, Mark Crovella and Aaron Mueller 56088

NE-MultiSpeech: A Multilingual Speech Corpus and ASR Benchmark for Northeast Indian Languages
Badal Nyalang and Angana Borah 56109

FactRot: Auditing Typed Fact Survival Under Iterative LLM Rewriting
Vladislav Kulikov, Mikhail Mozikov, Aleksei Stepin, Boris Tolstokulakov, Artyom Kabanov, Nasonov Ivan, Nikita Glazkov and Ilya Makarov 56120

Probing Stability–Plasticity Tradeoffs in Agent Memory through Cognitive Experimental Paradigms
Jiaqi Ding and Guorong Wu 56130

A Language Model from 1913: Pretraining on Historical Text
Xiaoxi Luo, Zachary Shinnick, Niclas Griesshaber, Yixuan Wang, Junchi Yu, Freda Shi, Philip Torr and Yao Lu 56150

SATURN: Symbolic Spatial Reasoning for Multi-Perspective Grounding
Danial Kamali, Tanawan Premsri, Shreya Rajpal, Amir Zadeh, Chuan Li and Parisa Kordjamshidi 56166

Left-Branching Transformers Excel at Right-Branching Languages: Data Shapes Word Order Preferences in Language Models
Varvara Arzt, Allan Hanbury and Terra Blevins 56198

Contrast Comes at a Cost: Prompting Large Language Models for Perception-Centered Gendered Style Transfer
Hongyu Chen, Neele Falk, Michael Roth and Agnieszka Falenska 56220

Steer, Don't Solve: Training Small Critic Models for Large Code Agents
Shubham Gandhi, Yiqing Xie, Atharva Naik, Ruichen Zhu and Carolyn Rose 56259

TDGNet: Hallucination Detection in Diffusion Language Models via Temporal Dynamic Graphs
Arshia Hemmat, Philip Torr, Yongqiang Chen and Junchi Yu 56277

Kronecker Factorization Improves Efficiency and Interpretability of Sparse Autoencoders
Vadim Kurochkin, Yaroslav Aksenov, Daniil Laptev, Daniil Gavrilov and Nikita Balagansky 56302

Reasoning Beyond English: Separating Capability and Language in Cross-Lingual Reasoning Model Adaptation
Jan Eliasz, Mikołaj Langner, Dzmitry Pihulski, Konrad Wojtasik, Przemyslaw Kazienko, Jan Kocon and Teddy Ferdinan 56324

MedFrameQA: A Multi-Image Medical VQA Benchmark for Clinical Reasoning
Suhao Yu, Haojin Wang, Juncheng Wu, Luyang Luo, Jingshen Wang, Cihang Xie, Pranav Rajpurkar, Carl Yang, Yang Yang, Kang Wang, Yannan Yu and Yuyin Zhou 56345

A Contextual Measure of Ambiguity in Prescriptive Texts
Sally Dong, Luke Thorburn, Andrew Konya and Michiel A. Bakker 56379

Consistency-Calibrated Advantage Estimation for Label-Free Reinforcement Learning
Meng Yang, Yongle Zhao, Zhiyang Zhang, Xiaorui Jiang, Peirou Liang, Furui Chen, Yin Xie and Ziyong Feng 56413

Universal-Guided Grammar Induction
Yuan Gao and Weiwei Sun 56426

STEEREVAL: Inference-time Interventions Strengthen Multilingual Generalization in Neural Summarization Metrics
Silvia Casola, Ryan Soh-Eun Shim, Felicia Körner, Yuchen Mao and Barbara Plank 56439

PRISM: Recovering Instruction Sets from Language Model Activations
Gilad Gressel, Rahul Pankajakshan, Julia Diament, Efim Hudis, Krishnashree Achuthan and Yisroel Mirsky 56467

ParaStudent: Closing the Sim2Real Gap in User Simulators for AI Tutor Evaluation
Rose Niousha, Mihran Miroyan, Abigail O'Neill, Joseph E. Gonzalez, Gireeja Ranade, John DeNero and Narges Norouzi 56493

Which Forms of Caregiver Feedback Support Grammar Learning? A Reinforcement-Learning Study of Child-Like Language Models
Jing Liu, Marianne Schweitzer and Abdellah Fourtassi 56512

EvaLex: Automatic Evaluation of Lexical Competence in Language Models
Jaime Collado-Montañez, L. Alfonso Ureña and Arturo Montejo-Ráez 56526

Entity Extraction from EU acquis
Jakub Piskorski, Dominik Skotarczak and Océane Abrard 56540

HyDRA: Hyperbolic Damage-aware Recovery and Adaptation for Robust Multimodal Sentiment Analysis with Incomplete Data
Yun Luo, Zejiang He, Zhen Huang, Zhiliang Tian, Yanzhou Ma, Menglong Lu, Jingyuan Huang, Shanshan Liu, Faying Wu and Dongsheng Li 56575

OS-Marathon: Benchmarking Computer-Use Agents on Vast-Horizon, Repetitive Tasks
Jing Wu, Wenjie Ai, Daphne Barretto, Yiye Chen, Qingyu Chen, Yuhang He, Pranit Chawla, Nicholas Gydé, Yanan Jian and Vibhav Vineet 56604

HSEARCH-R1: Learning Compact Value Models for LLM-guided Hierarchical Search via RL
Nilesh Gupta, Nasib Ullah, Jinbin Zhang, Rohit Babbar and Inderjit S Dhillon 56623

Prompt2Box: Improving LLM Weakness Discovery and Specificity Estimation by Uncovering Entailment Structure among Prompts
Neeladri Bhuiya, Shib Sankar Dasgupta, Andrew McCallum and Haw-Shiuan Chang 56643

Examining Reasoning LLMs-as-Judges in Non-Verifiable LLM Post-Training
Yixin Liu, Yue Yu, DiJia Su, Sid Wang, Xuewei Wang, Song Jiang, Bo Liu, Arman Cohan, Yuandong Tian and Zhengxing Chen 56671

POISS: A Large-Scale Multilingual Dataset for Point-of-Interest Search
Nicola Maritan, Alessandro Moschitti, Federico Borazio, Xiaokun Zhou and Zhengwei Bai 56705

From Cause to Effect: Generating Intermediate Event Chains in Videos
Ridouane Ghermi, Vicky Kalogeiton and Ivan Laptev 56724

When Does Reasoning Help in Machine Translation? A Hierarchical Analysis of LRM Reasoning Traces
Yuxiang Liu, Jiaming Luo, Eleftheria Briakou and Colin Cherry 56747

SABER-Math: An Automated Reranking Benchmark for Mathematical Information Retrieval
Nikolay Georgiev, Maria Drencheva, Kseniia Ibragimova, Ivo Petrov, Dimitar Iliev Dimitrov and Martin Vechev 56772

MUDIDI: A Two-Stage Framework for Multilingual Dictionary Digitization with Language Models
David Samuel Setiawan, Temuulen Khishigsuren, Milind Agarwal, Pagnarith Pit, Aso Mahmudi and Ekaterina Vylomova 56806

M^3EC: A Multimodal Multidocument Benchmark for Event Extraction and Event Coreference Resolution
Zaber Ibn Abdul Hakim, Najibul Haque Sarker, Aafiya Hussain, Hani Alomari, Alvi Md Ishmam, Ali Asgarov, Chia-Wei Tang and Chris Thomas 56843

HEAR Who Said What: Unlocking Speaker-Attributed Reasoning via Counterfactual Voice Grounding
Dongwook Lee, Sangkwon Park, Eunwoo Song, Che Hyun Lee, Youngho Cho, Junho Kim, June Young Yi, Heeseung Kim and Sungroh Yoon 56905

Rethinking Language's Role in Efficient VLA for Autonomous Vehicles: Toward Smarter, Trustworthy Driving
Tongfei Guo and Lili Su 56933

Structured-Sparse Attention for Entity Tracking with Subquadratic Sequence Complexity
Hangyue Zhao, Paul Caillon, Erwan Fagnou and Alexandre Allauzen 56948

Harm Laundering in GPT Models: Evidence That Gender Discrimination Is Transformed Rather Than Reduced Across Safety-Trained Generations
Sarah Wyer, Sue Black and Noura Al Moubayed 56960

Exploring Sparse Autoencoders in Text-Based Causal Confounding Adjustment
Mian Zhong, Katherine A. Keith and Anjalie Field 56985

Steering Geometry: Validating Human Value Geometry in LLM Steering Space
Mohammad Mahdi Abootorabi, Armin Saghafian, Ali Bazshoushtari, Hamid Rezaei, EunJeong Hwang, Vered Shwartz, Parvin Mousavi and Purang Abolmaesumi 57010

Two Centuries of Sexism in British Parliament: A Computational Analysis of Women's Representation in the Hansard Corpus
Mohammad Omar Khursheed, Mandira Sawkar and Ashiqur R. KhudaBukhsh 57034

DARTS: Decoder-Aware Representation Tuning via Surgery for Model Merging
Aaryan Ajay Sharma, Sai Nishanth Padala and Seganrasan Subramanian 57052

MetaCaster: Meta-Harness-Optimized Agent for End-to-End Few-Shot Learning of Lightweight Time Series Forecasters
ChengAo Shen, Wenchao Yu, Fangyu Wu, Dongjin Song, Hanghang Tong, Dongsheng Luo, Wei Cheng, Haifeng Chen and Jingchao Ni 57066

NEST: Narrative Event Structures in Time for Long Video Understanding
Ali Asgarov, Kaushik Narasimhan, Najibul Haque Sarker, Hani Alomari, Chia-Wei Tang, Anushka Sivakumar, Zaber Ibn Abdul Hakim, Shaurya Mallampati and Chris Thomas 57094

InfoEdit: Probing Global Layout Reasoning in Infographic Editing
Cheng Yang, Chufan Shi, Huijuan Wang, Bo Shui, Yaokang Wu, Muzi Tao, Yibo Yan, Xuezhe Ma and Taylor Berg-Kirkpatrick 57130

AgentKV: Phase-Aware KV Eviction for Agentic LLMs
Taowen Tony Liu, Jeffrey T. H. Wong, Can Xiao, Bowen Yang, Hao Mark Chen and Yiren Zhao 57169

Chronoberg: A 250-Year Diachronic Corpus with Temporal Affective Lexicons for Language Model Analysis
Niharika Hegde, Subarnaduti Paul, Lars-Joel Frey, Manuel Brack, Kristian Kersting, Martin Mundt and Patrick Schramowski 57185

WhisTLE: Deeply Supervised, Text-Only Domain Adaptation for Small Pretrained Speech Recognition Transformers
Akshat Pandey, Karun Kumar and Raphael Tang 57220

Geometric-disentanglement Unlearning
Duo Zhou, Yuji Zhang, Tianxin Wei, Ruizhong Qiu, Ke Yang, Xiao Lin, Cheng Qian, Jingrui He, Hanghang Tong, ChengXiang Zhai, Heng Ji and Huan Zhang 57231

Robust and Efficient Guardrails with Latent Reasoning
Siddharth Sai, Xiaofei Wen and Muhao Chen 57256

UReason: Benchmarking Reasoning-to-Generation Alignment in Unified Multimodal Models
Cheng Yang, Chufan Shi, Bo Shui, Yaokang Wu, Muzi Tao, Huijuan Wang, Ivan Yee Lee, Yong Liu, Xuezhe Ma and Taylor Berg-Kirkpatrick 57269

REFRAME: A Test Suite for the Knowledge–Deployment Gap in LLMs
Elahe Rahimi, François Roewer-Després, Hassan Sajjad, Md Mahbub Alam, Sher Badshah, Elham Dolatabadi and Frank Rudzicz 57310

AegisCodeGen: Toward Copyright-Aware Code Generation with Runnability-Guided Fusion
Rui Chu, Nicole Meng, Tanusree Das Tithy, Thomas Cao, Faysal Hossain Shezan and Yingjie Lao 57326

PetQA: Benchmarking Veterinary Knowledge and Clinical Reasoning
Taegyun Kim, Youngwook Ham, Jungwook Rhim, Ju-Hyun An, Sungkyu Park and Kunwoo Park 57338

Beyond "To whom it may concern": Tailoring Machine Translation to Audience and Intent
Raphaël Merx, Ekaterina Vylomova and Trevor Cohn 57366

Visual Framing for News Stance Detection via Image Generation
Dahyun Lee, Jiyoung Han and Kunwoo Park 57393

Who Wrote the Book? Detecting and Attributing LLM Ghostwriters
Anudeex Shetty, Qiongkai Xu, Olga Ohrimenko and Jey Han Lau 57419

Understanding Quantization of Optimizer States in LLM Pre-training: Dynamics of State Staleness and Effectiveness of State Resets
Kristi Topollai and Anna Ewa Choromanska 57448

Prompt-Induced Recovery Behavior Decomposition for Black-Box Pre-training Data Detection
Xin Fan, Miyamoto Ryoto, Fan Mo, Chongxian Chen, Tsuneo Matsumoto, Fuyuko Kido and Hayato Yamana 57471

Is Agent Code Less Maintainable Than Human Code?
Shaswat P Patel, Betty Li Hou, Arun Purohit, Kai Xu, Jane Pan, He He and Valerie Chen . 57491

Can LLMs Use Linguistic Uncertainty Markers to Reliably Reflect Intrinsic Confidence?
Gabrielle Kaili-May Liu and Arman Cohan 57515

Typed Relation Path Geometry in Sentence Embedding Spaces
Mingyue Liu and Ziyang Xing 57548

Hidden Configuration Effects on Fairness in LLM-Based Multi-Agent Decision Systems
Hamed Loghmani, Ninareh Mehrabi, Homa Hosseinmardi and Amir Ghasemian 57565

Finetune Once: Decoupling General & Domain Learning with Dynamic Anchor Annealing
Yang Tang, Ruijie Liu, Yifan Wang, Shiyu Li, Peiming Li and Zheng Wei 57590

Sycophancy Is Not One Thing: Causal Separation of Sycophantic Behaviors in LLMs
Daniel Vennemeyer, Phan Anh Duong, Tiffany Zhan and Tianyu Jiang 57608

To Trust or Not to Trust: Retrieval-Augmented Fact Checking in Speech
Debajyoti Mazumder, Mamta Mamta and Abhirama Subramanyam Penamakuri 57632

Recontextualizing Famous Quotes for Brand Slogan Generation
Ziao Yang, Zizhang Chen, Lei Zhang and Hongfu Liu 57651

Collaborative Reasoning with Reranked Multi-facet Retrieval for Legal Judgment Prediction
Wei Cai, Zihan Mei, Haojun Huang and Bang Wang 57666

Text Capability Loss in Vision-Language Adaptation: An Attention-Sink Diagnosis
Minsik Choi, Geewook Kim and Young Geun Kim 57684

Discerning Culturally-Aware Hate Framings for Hate Speech Identification in Memes
Rakshitha Rao Ailneni and Sanda M. Harabagiu 57713

Multilingual Chain of Thought Compression via Cross-lingual Entropy Gap
Jiarui Wan, Yufeng Chen, Songming Zhang, Rui Qi, Xue Zhang, Jinan Xu and Kaiyu Huang 57744

Broken Symmetry in LLM Refusal: Answer Release Is More Local Than Refusal Restoration
Yiqi Liu, Yang Wang, Songxin Wang, Chenghao Xiao and Chenghua Lin 57765

TTS^2: Test-Time Scaling for Flow-Based Text-to-Speech via Multi-stage Reward Modeling on Posterior Mean Projections
Tung Luong Nguyen, Tuan Duy Nguyen Le, Nhat Hong Tran, Nhat Minh Le and Van Hai Do 57786

ManimAgent: Self-Evolving Multimodal Agents for Visual Education
Wenjia Jiang, Zongyuan Cai, Yuanhang Shao, Chenru Wang, Boyan Han, Zhixue Song, Keyu Chen, Shengwei An, Xu Yang and Zhou Yang 57798

Cross-Preference Learning for Sentence-Level and Context-Aware Machine Translation
Ying Li, Xinglin Lyu, Junhui Li, Jinlong Yang, Hengchao Shang, Min Zhang, Shimin Tao and Daimeng Wei 57823

FlipTS: Failure-to-Success Mathematical Self-Refinement via Trajectory-Aware Sampling in End-to-End Online RL
Yaqian Hao, Yifan Hao, Yankai Liu, Fangning Chao, Zhaojun Cui, Haiyu Zhang, Huan Bai, Fanyu Meng, Chao Deng and Junlan Feng 57845

Scoped Semantic Invariance as a Foundation for LLM Reliability
Rosario Uceda-Sosa, Maria Chang, Karthikeyan Natesan Ramamurthy and Moninder Singh 57863

Investigating and Alleviating Harm Amplification in LLM Interactions
Ruohao Guo, Wei Xu and Alan Ritter 57885

Don't Peek at the Answer: Outcome-Masked Group Relative Policy Optimization for Label-Free RLVR
Yongshi Ye, Liang Zhang, Yidong Chen, Xiaodong Shi and Biao Fu 57905

FrankenReport: Early Exiting in Long-Form Generation Using Expected Value of Computation
Zhengping Jiang, Gonzalo Ramos, Jina Suh, Shiqian Rachel NG, Elias Stengel-Eskin, Justin Svegliato, Benjamin Van Durme, Andy Huntington and Sam Thomson 57929

PEARL: Front-Loading Relational Chains for Multi-Hop Table Retrieval
Subeen Ho, Hyeongu Kang, SeongKu Kang and Susik Yoon 57952

PAMT: Process-Aligned Reinforcement Learning for Multi-Domain Machine Translation
Yongshi Ye, Biao Fu, Chongxuan Huang, Yidong Chen and Xiaodong Shi 57977

Not All Memories Are Equal: Hierarchical Collaborative Memory for Validity-Aware Retrieval in LLM Agents
Yufei Shi, Rujing Yao, Ang Li, Yang Wu, Zhuoren Jiang and Xiaozhong Liu 58002

Visual Preference Optimization with Rubric Rewards
Ya-Qi Yu, Fangyu Hong, Xiangyang Qu, Hao Wang, Gaojie Wu, Qiaoyu Luo, Nuo Xu, Huixin Wang, Wuheng Xu, Yongxin Liao, Zihao Chen, Haonan Li, Ziming Li, Dezhi Peng, Minghui Liao, Jihao Wu, Haoyu Ren and Dandan Tu 58016

TTML: Test-Time Memory Learning for Label-Free RLVR
Yongshi Ye, Biao Fu, Xinjie Chen, Minpeng Liao and Xiaodong Shi 58038

Reinforcement Learning with Robust Rubric Rewards
Ya-Qi Yu, Hao Wang, Fangyu Hong, Xiangyang Qu, Gaojie Wu, Qiaoyu Luo, Nuo Xu, Huixin Wang, Wuheng Xu, Yongxin Liao, Zihao Chen, Haonan Li, Ziming Li, Dezhi Peng, Minghui Liao, Jihao Wu, Haoyu Ren and Dandan Tu 58060

From Chain to Mesh: Evolving Robust Safety Alignment via Neuron-Circuit Topology
Guan Wang, Xuehai Tang, Biyu Zhou, Wenjie Xiao, Yuanshu Zhao, Lijia Lv, Jizhong Han and Songlin Hu 58079

BalancedBio: Mitigating the Alignment Tax in Biomedical LLMs via Gradient Orthogonality and GRPO
Wentao Wu, Hanmeng Zhong, Linqing Chen and Peng Zhou 58101

From Shallow Alignment to Source Correction: Robust Safety Alignment for LLMs
Guan Wang, Xuehai Tang, Biyu Zhou, Jizhong Han and Songlin Hu 58116

How to Solve Noise and Hard Samples: Filtering and Difficulty-adaptive Augmentation for Fine-grained Negative Emotion Classification
Lianxi Wang, Siting Lin, Chao Liang, Yuying Li, Xinfeng Liao and Nankai Lin 58138

Investigating Length Bias and Robustness in LVLMs for Multiple-Choice Question Answering
Md. Atabuzzaman, Hani Alomari and Chris Thomas 58157

TRUST-SQL: Tool-Integrated Multi-Turn Reinforcement Learning for Text-to-SQL over Unknown Schemas
Ai Jian, Xiaoyun Zhang, Eryu Guo, Wanrou Du, Jingqing Ruan, Haojin Yang, Jiangbo Pei, Weipeng Zhang, Ke Zeng and Xunliang Cai 58177

Reliability Challenges in Diffusion Vision–Language Models
Md. Atabuzzaman and Chris Thomas 58201

VAA-CSEC: Vote-guided Advantage Allocation for Chinese Semantic Error Correction
Yitong Han, Nankai Lin, Juan Luo, Hongyan Wu, Lianxi Wang and Shengyi Jiang 58217

DocHop-QA: Towards Multi-Hop Reasoning over Multimodal Document Collections
Jiwon Park, Seohyun Pyeon, Jinwoo Kim, Rina Carines Cabral, Zhenyuan He, Yihao Ding and Caren Han 58233

Neuro-Symbolic Reasoning with External Inference Tools Mitigates Bias from Sub-Formula Level Logical Structure
Ryo Hasegawa, Yusuke Sakai, Hidetaka Kamigaito and Taro Watanabe 58262

SkillAlign: Aligning Skill Interfaces for LLM-based Agents
Shuo Ren, Xiaomian Kang and Jiajun Zhang 58285

Alignment or Formatting? Dissecting What On-Policy DPO Actually Learns
Keyu Wen, Yang Song, Lijun Yang and Rui Lin 58301

Diversity is Not Ambiguity: Toward Accurate and Efficient Ambiguity Detection for Open-Domain QA
Jiwon Lee, Yong-chan Park, Jungin Hong and U Kang 58310

TriVTab: Fusing Triple-Based Structured Reasoning and Visual Reasoning for Multimodal Table Question Answering
Zhuo Wen, Yang Lin, Weifan Zhao, Ji Zhang and Jun Gao 58334

Silence the Judge: Reinforcement Learning with Self-Verifier via Latent Geometric Clustering
Nonghai Zhang, Weitao Ma, Zhanyu Ma, Jun Xu, Jiuchong Gao, Jinghua Hao, Renqing He and Jingwen Xu 58355

Overfitting Mitigation via Singular Value Decomposition in Minimum Bayes Risk Decoding
Riza Setiawan Soetedjo, Yusuke Sakai, Hidetaka Kamigaito, Katsuhiko Hayashi and Taro Watanabe 58374

Security and Privacy Prompts in the Wild: What Users Ask LLMs and How LLMs Respond
Hobin Kim, Xiaoyuan Wu, Omer Akgul, Lujo Bauer and Nicolas Christin 58402

Multimodal Collaborative Debate for Zero-Shot Time Series Reasoning
Patara Trirat, Jin Myung Kwak, Jay Heo, Heejun Lee and Sung Ju Hwang 58432

ExTax: Explainable Disinformation Detection via Persuasion, Emotion, and Narrative Role Taxonomies
Shang Luo, Yingguang Yang, Zhenchen Sun, Yang Liu, Bin Chong, Jingru Chen, Yancheng Chen, Jiayu Liang, Xu Kefu, Hao Peng and Philip S. Yu 58457

MemoryRewardBench: Benchmarking Reward Models for Long-Term Memory Management in Large Language Models
Zecheng Tang, Baibei Ji, Ruoxi Sun, Haitian Wang, Juntao Li and Min Zhang 58484

Mechanistic Diagnostics of Spatial Lexical Bias in Multimodal Large Language Model Spatial Reasoning
Chuang Ma, Qianying Liu, Tomoyuki Obuchi, Fei Cheng, Wang Yang, Sudong Cai, Shuyuan Zheng, Akiko Aizawa and Sadao Kurohashi 58517

JAMER: Project-Level Code Framework Dataset and Benchmark on Professional Game Engines
Jianwen Sun, Chuanhao Li, Zizhen Li, Yukang Feng, Fanrui Zhang, Yifei Huang, Yu Dai and Kaipeng Zhang 58544

One Refusal Isn't Enough: Multi-Turn Jailbreak Attacks on Large Language Models
Pagnarasmey Pit, Xingjun Ma, James Bailey and Eduard Hovy 58571

Validity-Aware Assessment of LLM Value Orientation via Subset Item Response Theory
Qinlin Zhao, Xiaoyuan Yi, Jing Yao, Muhua Huang, Defu Lian and Xing Xie 58592

Evaluating LLM Watermarking for Clinical EHR: Detectability, Style Shifts, and Helpfulness
Rochana Prih Hastuti, Rian Adam Rajagede, Mengxin Zheng and Qian Lou 58616

ACE: Adapter Consolidation across Experts for Parameter-Efficient Fine-Tuning of MoE LLMs
Ahin Lee, Sehyun Yun, Joonha Park and Taesik Gong 58640

ParaSpeechRAG: A Benchmark for Robust Speech Retrieval under Paralinguistic and Acoustic Perturbations
Alexandre Costa Ferro Filho, Marcos Paulo C M Queiroz, Matheus Franco Medeiros, Gabriel Orlow De Oliveira Sene, Murilo Alvares Silva, Diogo Fernandes Costa Silva and Arlindo Rodrigues Galvão Filho 58663

Thunder-DeID+: Boundary-Aware Korean De-identification from Court Judgments to Broader Sensitive Documents
Gyuseong Lee, Gyeongje Cho, Jongmin Kim, Heejin Kim, Yeonkyoung So, Hyunggeun Jeon, Hyunji M. Park, Sangho Kim, Yoonhee Park, Youngbin Pyo and Jaejin Lee 58675

GeoChallenge-90K: A Multi-Answer Multiple-Choice Benchmark for Diagram-Grounded Geometric Reasoning
Yushun Zhang, Weiping Fu, Zesheng Yang, Bo Zhao, Lingling Zhang, Jian Zhang, Yumeng Fu, Jiaxing Huang and Jun Liu 58710

Responsibility-Aware Scheduling for Selective Rationalization
Dengchao Zhang, Yunxiao Zhao, Zhichao Yan, Yue Fan, Shaoru Guo and Ru Li 58731

IntroConformal: Conformal Factuality Guarantees for Large Vision-Language Models via Introspective Signals
Md. Atabuzzaman, Christian Alexander and Chris Thomas 58747

MMLongEmbed: Benchmarking Multimodal Embedding Models in Long-Context Scenarios
Haitian Wang, Ruoxi Sun, Quantong Qiu, Juntao Li, Junhui Li, Hua Chen, Jinxiong Chang and Min Zhang 58765

Balancing Epistemic Value and Social Safety: PACR-Pro for False-Consensus Intervention in One Night Ultimate Werewolf
Yonghui Wang, Zhiliang Tian, Bo Liu, Jiarui Tang, Wanlong Yu and Shilong Pan 58786

CMPA-SLU: A Benchmark for Chinese Mobile Personal Assistant Spoken Language Understanding
Di Wu, Zhongqiu Li, Zetao Zhu, Sishi Xiong, Ruiyu Fang and Shuangyong Song 58811

ConflictScore: Identifying and Measuring How Language Models Handle Conflicting Evidence
Siyi Liu, Aaron Halfaker, Dan Roth and Patrick Xia 58825

Second language input density can induce age of acquisition effects in language models
Simon Jerome Han and James McClelland 58846

Training-Free Chunk-Level Guided Generation for Mathematical Reasoning
Atoosa Chegini and Soheil Feizi 58859

Intra-Prompt Parallel Decoding for Common-Context Question Answering
Theodore Glavas, Nikhita Vedula, Dushyanta Dhyani, Antonios Valkanas, Yilun Zhu and Shervin Malmasi 58871

V-CoLA: Vision Token Compression with Linear Attention
Hao Jiang, Yiru Mao, Tianpeng Bu, Hao Zhou, Hongtao Duan, Wang Jing, Bowen Xu, Xin Chen, Lulu hu, Bin Yang, Yongliang Tao and Minying Zhang 58891

TRACE: Table–Text Relationalization and Completion of Evidence for Coverage-Sensitive QA
Sunho Cha, Sungho Park and Wook-Shin Han 58907

Where Do Vision-Language Models Fail on Abstract Reasoning? An Information Processing Analysis of ARC
Xuan Zhou, Xuefeng Wei, Zhi Qu, Yusuke Sakai, Hidetaka Kamigaito and Taro Watanabe 58927

DySem: Uncovering Dynamic Semantic Components of Large Language Models for Calculating Semantic Textual Similarity
Kaijie Zheng, Weiqin Wang, Yile Wang and Hui Huang 58948

AMNESIA: A Large-Scale Medical Unlearning Benchmark Suite with Disease-Informed Analysis
Saeedeh Davoudi, Reihaneh Iranmanesh, Ophir Frieder and Nazli Goharian 58968

Accountable AI with Grounded, Faithful, Consistent, Actionable Rationales: A Case Study in Clinical Trial Matching with VERDICT
Zikai Zhou, Yufei Jin, Yilin Xu, Yu-Chiang Wang, Chieh-Ju Chao and Monica Lam 58987

FineVerify: Scaling Test-Time Compute with Fine-Grained Self-Verification for Agentic Search
James Xu Zhao, Hui Chen, Bryan Hooi and See-Kiong Ng 59033

Lazy Grounding: Attacking Search Agents with Factual Evidence
Yulin Zhang, Yukun Huang, Sanxing Chen, Tianyi Lin, Ziang Yang, Xunjian Yin and Bhuwan Dhingra 59063

Residual Sparsification via Output Importance for Compressing Mixture-of-Experts LLMs
Seungwoo Jung, Dohyeok Kwon, Seungmin Cha, Junseok Lee, Yeonho Yoo, Chuck Yoo and Gyeongsik Yang 59078

Beyond Token Noise: Discourse-Level Rewrites Expose Brittleness in QA LLMs
Mousa Arraf, Ido Guy and Kira Radinsky 59098

Dual Controller Architecture to Enable Adherence of LLM-based Counselors to Session- and Discourse-level Aspects of Motivational Interviewing Theory
Yermakhan Magzym, April Idalski Carcone, Elizabeth Towner, Christopher Mann M.D., M. Safwan Badr and Alexander Kotov 59120

Beyond Analysis: Training Language Models with Internal Mechanistic Feedback
Hanlin Xue, Gang Zhang, Xiangyuan Wang, Yunhao Bai, Yulong Wan, Xiang Li, Jiaqi Zhou, Ruochong Xiong and Junfei Liu 59164

Bridge Routing Heads: Where Multilingual Multi-hop Reasoning Lives in LLMs
Seunghan Kim, Minyeong Choe, Hyunil Kim and Haehyun Cho 59179

There Is More to Refusal in Large Language Models than a Single Direction
Faaiz Joad, Majd Hawasly, Sabri Boughorbel, Nadir Durrani and Husrev Taha Sencar 59196

AI Alignment through a Game-theoretic Lens: A Survey
Yanan Cai, Zhongrui Zhao, Zhigang Lu, Ickjai Lee, Wei Emma Zhang, Jason Xue, Yihong Zhang, Shuchao Pang and Wei Xiang 59233

Who Pays More for Safety? Measuring the Disparate Cost of Safety Alignment across Languages
Chanwoong Yoon, Jungsoo Park and Alan Ritter 59254

SCOPE: Stepwise Calibration of Reasoning Confidence with Outcome-Grounded Prefix Evidence
Huaqin Zhao, Yifan Zhou, Bolun Sun, Chao Cao, Yiwei Li, Zhaojun Ding, Lu Zhang, Tianming Liu and Jin Lu 59284

Deconstructing Spatial Complexity: Hierarchical Decomposition for LLM Spatial Reasoning
Yi Wang, Haojie Lu, Chen Li and Sihong Xie 59298

Q-TIE: A Lightweight and Generalizable Re-ranking Framework for Temporal Information Retrieval
Soyeon Kim, HyunJin Kim, JinYeong Bak and Steven Euijong Whang 59320

Judging LLM-as-a-Judge: Concerning Rubric Artifacts in LLM-based Automated Text Generation Evaluation
Anshul Bagaria, Sowmya S Sundaram, Gokul S Krishnan and Balaraman Ravindran 59341

SciRerankBench: Benchmarking Rerankers Towards Scientific Retrieval-Augmented Generation with LLMs
Haotian Chen, Qingqing Long, Meng Xiao, Xiao Luo, Wei Ju, Yuanchun Zhou, Xuezhi Wang and Hengshu Zhu 59353

Graph-of-Skills: Dependency-Aware Structural Retrieval for Massive Agent Skills
Dawei Liu, Zongxia Li, Hongyang Du, Xiyang Wu, Shihang Gui, Yongbei Kuang and Lichao Sun 59381

VEX-Bench: Benchmarking LLM Agents for Assessing Exploitability of Software Supply Chain Vulnerabilities
Jiahao Shi, Edward Tsien, Yifeng Di, Hongjiao Zhang, Yuan Tang, Ronit Dey, Ilona Shishov, Gal Netanel, Zvi Grinberg, Vladimir Belousov, Bat-Zion Rotman, Ilan Pinto and Tianyi Zhang 59409

SurveyLens: A Discipline-Aware Benchmark for Automatic Survey Generation
Beichen Guo, Zhiyuan Wen, Jia GU, Haochen Shi, Jian Wang, Senzhang Wang, Haoyang LI, Ruosong Yang and Shuaiqi Liu 59426

MULTI3IR: A Benchmark for Multi-perspective Multi-domain Multi-modal Information Retrieval
Seokwon Song, Sohyeon Kim and Gunhee Kim 59459

DuplexGen: Adaptive Synthesis of Human-AI Turn-Taking Dialogues
Takyoung Kim, Kang-wook Kim, Tony Woo, Julia Hirschberg, Gunhee Kim and Dilek Hakkani-Tür 59479

Idioms Understood, Yet Translated Literally: Diagnosing Literal Translation Bias in Multilingual LLMs
Yanming Sun, Runzhe Zhan, Yu Chao, Jingkun Ma, Shuo Wang, Lidia S. Chao and Derek F. Wong 59506

GUI-Primitives: Diagnosing Spatial Reasoning Failures in Vision-Language GUI Grounding
Md Abrar Jahin and Md Rizwan Parvez 59530

Rethinking Molecular Text Representations for LLMs: An Empirical Study
Arun Raja, Garrett M Morris and Kian Ming A. Chai 59565

Controllable Image Captioning with Prompt-Conditioned Scene Rewards
Jongyeop Hyun, Taeyoung Kim and Hyounghun Kim 59590

TAP: Policy Interfaces for Diffusion Language Models
Preeti, Ankita Kushwaha, Kiran Ravish and Pawan Kumar 59616

Vision-Braille: A Curriculum Learning Toolkit and Braille-Chinese Corpus for Braille Translation
Alan Wu, Ye Yuan, Zhiping Xiao and Ming Zhang 59658

Localizing Cross-Lingual Generalization in Parameter Space
Yeoreum Lee, Jaeseong Lee and Seung-Won Hwang 59673

TRACE: Temporal Retrieval with Anchored and Convergent Evidence for Long-Horizon Video Understanding
Pengyiang Liu, Junbo Niu, Xiaoyang Hu, Zhongyue Shi, Zitian Wang, Linjiang Huang and Si Liu 59693

DIPLOMAT: Dialogue-Span-Aware Direct Preference Optimization for Polite Persuasive Workplace Negotiation Dialogues
Bibhuti Jha, Rishikant Chigrupaatii, Priyanshu Priya and Asif Ekbal 59709

PRAGMA: Evaluating Personalized Guidance with Memory Alignment in Lifelong Conversations
Hyojeong Yu, Hyukhun Koh, Minsung Kim, Yunah Jang and Kyomin Jung 59750

Small Reward Models via Backward Inference
Yike Wang, Faeze Brahman, Shangbin Feng, Teng Xiao, Hannaneh Hajishirzi and Yulia Tsvetkov 59776

MARD: A Multi-Agent Reinforced Debiasing Framework for Large Language Models
Xiaoqing Cheng, Tao Liu, Tian Lan, Jinwang Song, Chen Chen, Hongying Zan, Wen Zhang and Huajun Chen 59812

HarmoLoRA: Commit-Stage Update Admission in Low-Rank Adaptation for Continual Speech Recognition
Ye-Eun Ko, Ji-Eun Choi, Dong-Hyun Kim and Joon-Hyuk Chang 59824

From Anonymous Speaker Seeds to Consistent Characters: Zero-Shot Manga Speaker Attribution
Yanhao Chen, Wataru Shimoda, Kota Yamaguchi and Hideki Nakayama 59842

Dango: A Strictly L1-Only Large Language Model for Studying Second Language Acquisition
Shiho Matta, Yin Jou Huang, Fei Cheng, Takashi Kodama, Hirokazu Kiyomaru and Yugo Murawaki 59856

Teach the Magnitude, Not the Direction: Verifier-Anchored Credit Assignment for Multi-Turn Multi-Step LLM Agents
Zechuan Wang, Siyuan Lu, Hongxuan Zhang, Linjian Mo, Chenyi Zhuang and Leilei Gan 59877

ParsBench: A Benchmark for Evaluating Fine-Grained Culturally Grounded Commonsense Reasoning in Persian
Mehrnoush Alizade, Farbod Davoodi and Suman Kalyan Maity 59892

No Matter Who You Are: Addressivity and Sycophancy in Large Language Models
Seungjong Sun, Hyunmin Jun, Minsik Yoon, Eungu Lee, Choongwon Kang, Young Seok Woo and Jang Hyun Kim 59914

Succinctness, Not Expressivity, Predicts Syntactic Competence
Shangao Li, Jinghan Wu, Chang Lu and Yizheng Zhao 59940

Extracting Small Translation Specialists from LLMs by Aggressively Pruning Experts
Liu Martin, Lucas Bandarkar and Nanyun Peng 59977

sk-bench: A Native-First Benchmark for Evaluating Large Language Models in Slovak
Marek Šuppa, Ivan Vykopal, Andrej Ridzik, Kristián Sopkovič, Natália Kňažeková, Jaroslav Kopcan, Miroslav Blstak, Viktória Ondrejová, Daniel Hládek, Michal Gregor, Martin Tamajka and Marian Simko 60008

TriAlign: Towards Universal Truth Consistency in Personalized LLM Alignment
Thi-Nhung Nguyen, Linhao Luo, Rollin Omari, Junae Kim, Thuy-Trang Vu and Dinh Phung60056

You Do Not Fully Utilize Transformer's Representation Capacity
Gleb Gerasimov, Yaroslav Aksenov, Nikita Balagansky, Viacheslav Sinii and Daniil Gavrilov 60080

Auditing Multi-Agent LLM Reasoning Trees Outperforms Majority Vote and LLM-as-Judge
Wei Yang, Shixuan Li, Gengshuo Liu, Heng Ping, Peiyu Zhang, Paul Bogdan and Jesse Thomason 60101

SeDeM: Selective Decompression of Hidden-State Memories for Long-Context Question Answering
Maryam Haghifam, Jason Cong and Yizhou Sun 60119

Beneath the Surface of Chains-of-Thought: A Mechanistic Interpretation of Reasoning Operations in LLMs
Seogyeong Jeong, Jaehui Hwang, Dongyoon Han, Geonmo Gu, Alice Oh and Taekyung Kim 60143

VisER: Visual Evidence and Reliance for Object Hallucination Detection in LVLMs
Afsaneh Hasanebrahimi, Hanxun Huang, Christopher Leckie and Sarah Erfani 60186

XYBench: Can LLMs Respond Pragmatically to Queries with Misconceptions?
Akhila Yerukola, Jena D. Hwang, Mingqian Zheng, Jenna Godsey, Hyunwoo Kim, Valentina Pyatkin, Jennifer Hu and Maarten Sap 60203

MultiModal Code-Switching: Interleaving Visual Objects into Language for Explicit Object-Level Alignment
Changhao Xiang, Shangyu Xing, Zhen Wu, Jianbing Zhang and Xinyu Dai 60235

What Matters in Linearizing Language Models? A Comparative Study of Architecture, Scale, and Task Adaptation
Patrick Haller, Jonas Golde and Alan Akbik 60257

Latent Task Embedding: Geometric Expansion for Single-Path Multi-Task PEFT
Junhyun Lee, Seoha Song and Dhananjaya Gowda 60275

Looks Right, Works Right: A Project-Level Benchmark for Multi-Screen Mobile App Generation
Fan Wu, Cuiyun Gao, Yiming Huang, Yang Xiao, Yujia Chen and Qing Liao 60294

Membrane: A Self-Evolving Contrastive Safety Memory for LLM Agent Defense
Minseok Choi, Seungbin Yang, Dongjin Kim, Subin Kim, Jungmin Son, Yunseung Lee, Jaegul Choo and Youngjun Kwak 60320

Retrieval Sensitivity to Identity Signals in Queries
Andrew Tang, Nicholas Deas, Kathleen McKeown and Vishal Misra 60355

PReMISE: Policy Rubrics as Measurement Specifications for LLM Judges
Swastik Roy, Rajkumar Pujari, Tharindu Kumarage, Charith Peris, Rahul Gupta, Anna Rumshisky, Pradeep Natarajan and Venkatesh Saligrama 60382

Shared Doubt: Zero-Shot Cross-Lingual Confidence Estimation for Language Models
Athina Kyriakou, Dennis Ulmer and Ivan Titov 60407

ALIGNBEAM : Inference-Time Alignment Transfer via Cross-Vocabulary Logit Mixing
Chirag Chawla, Pratinav Seth and Vinay Kumar Sankarapu 60438

Parts-of-Speech as Emergent Categories in SAE Latent Space
Alessandro Bondielli, Lucia Passaro, Serena Auriemma and Alessandro Lenci 60465

TEMPS: Temporal Sentence Embeddings for Temporal Information Retrieval
Mourad Hassani, Julien Romero, Amel Bouzeghoub and Jacquelinet 60482

ScienceArena: Benchmarking LLMs on Latest Scientific Olympiad Competitions
Guangxiang Zhao, Qilong Shi, Xusen Xiao, Liuwenpu, Yaoming Li, Linfeng Hao, Shuyang Hou, Zijian Guo, Xinrui Zhang, Yuntian Zhao, Zhengyang Wang, Wenrui Liu, Yuhan Wu, Tong Yang, Lin Sun and Xiangzheng Zhang 60506

Expectation, Backlash, Recovery, and Excitement: How Model Releases Shape Reddit Perceptions of Conversational AI Systems
Vahid Rahimzadeh, Yury Zhauniarovich and Savvas Zannettou 60524

HALvest-Contrastive: Retrieval-like Authorship Attribution with Patch-level Late Interaction
Francis Kulumba, Wissam Antoun, Guillaume Vimont, Laurent Romary and Florian Cafiero 60552

MeKi: Memory-based Expert Knowledge Injection for Efficient LLM Scaling
Ning Ding, Fangcheng Liu, Kyungrae Kim, Linji Hao, Kyenghun Lee, Hyeonmok Ko and Yehui Tang 60572

Probing Chemical Language Models: Effects of Pre-training and Fine-tuning
Anna Karnysheva, Dietrich Klakow and Ji-Ung Lee 60585

Cross-Corpus Depression Detection in Semi-Structured Clinical Interviews
Misha Sadeghi, Phuc Truong Loc Nguyen, Andreas Triantafyllopoulos, Mahdis Habibpour, Robert Richer, Lydia Helene Rupp, Lena Schindler-Gmelch, Marie Keinert, Malin Hager, Björn Schuller, Bernhard Egger, Matthias Berking and Bjoern Eskofier 60628

Language-Switching Triggers Take a Latent Detour Through Language Models
Francis Kulumba, Wissam Antoun, Théo Lasnier, Benoît Sagot and Djamé Seddah 60647

Annotators of a Feather FLOCKE Together: A Framework to Analyze Latent Group Patterns Based on Annotator Behavior
Giovanni Valer, Elisa Leonardelli, Camilla Casula, Sara Tonelli and Michele Lombardi . . . 60662

Memory Transfer Learning: How Memories are Transferred Across Domains in Coding Agents
Kangsan Kim, Minki Kang, Taeil Kim, Yanlai Yang, Mengye Ren and Sung Ju Hwang . . . 60677

Ouvia: A User-centered Framework for Measuring Usability of Speech Translation in Real-World Communication Scenarios
Giuseppe Attanasio, Beatrice Savoldi, Daniel Chechelnitsky, Matteo Negri, Marine Carpuat, Maarten Sap and Andre Martins 60698

REFOCUS: Training-Free Attention Sparsification via Uniform-Reference Filtering
Fengyuan Fan, Rena Wei Gao and Yanbei Jiang 60717

Can Activation Steering Capture Multidimensional Authorship Style?
Hieu Tran, Calvin Bao and Marine Carpuat 60731

Layer-wise Target Propagation: Efficient Component Attribution through Target Centric Propagation
Lasse M. Jantsch, Dong-Jae Koh, Seonghyeon Lee and Young-Kyoon Suh 60755

Disentangling Spurious Correlations: How IRM Handles Diverse Shortcuts in Text Classification
Paul Grunenwald, Benoit Crabbé, Guillaume Wisniewski and Pirmin Lemberger 60785

TurCOMET: A Morphologically-Aware Evaluation System for Document-Level Turkish-English Machine Translation
Atahan Uz, Hüseyin Emir Akdağ and Susan Üsküdarlı 60806

SPHERE: Automatic Music Upmixing via Audio Language Model Post-Training with Spatial Heuristics Rewards
Zixun Guo, Calvin Murdock, Sanjeel Parekh, W. Owen Brimijoin, Simon Dixon, Joshua Reiss and Ishwarya Ananthabhotla 60827

SimCRAFT: Distilling Remote Sensing Agents via Synthetic Trajectories and Contextual Retrieval-Augmented Fine-Tuning
Haoran Wang, Jing Yao, Xu Yang, Zeqing Wang, Yang Zhang, Pedram Ghamisi and Zhengchao Chen 60839

POLAR: Polarity-Aware and Anatomy-Conditioned Radiology Report Evaluation
Lukas Buess, Tri-Thien Nguyen, Xiaoliang Wang, Paula Andrea Perez-Toro, Tomas Arias-Vergara, Bernhard Kainz and Andreas Maier 60854

Pair Difficulty Matters: Rethinking Pairwise LLM-as-a-Judge Evaluation and Consistency
Bruno Brocai and Maria Becker 60873

Towards Multidisciplinary Citance Analysis: A Novel Approach for Intent, Polarity, and Semantics Classification
Georgios Pantelakis, Petros Stavropoulos and Haris Papageorgiou 60882

Berry-LR: Literature Retrieval for Automatic Scientific Survey Generation via Berrypicking Model
Yajiao Wang, Yang Li, Xiaotao Wang, Xun Feng, Yi Liu, Mengting Zhang, Xin Lin, Maodi Hu and Zhixiong Zhang 60908

Rethinking the Role of Tensor Decompositions in Post-Training LLM Compression
Artur Zagitov, Alexander Miasnikov, Maxim Krutikov, Artem Tsedenov, Vladimir Aletov, Gleb Molodtsov, Nail Bashirov and Aleksandr Beznosikov 60931

Deconstructing Memorization and Generalization: Dual-Level Calibration of Membership Inference Attacks for Large Vision-Language Models
Xiaofan Zheng, Xinghao Wang and Minnan Luo 60954

You Have a Face for Radio: Bias in Multimodal Speech Recognition Models
Maya K. Nachesa, Vlad Niculae and Vagrant Gautam 60970

Once a Tree, Always a Tree? Cross-lingual Transfer of Hierarchical Generalization in Language Models
Shin Yamaguchi, Masaya Taniguchi, Tatsuki Kuribayashi, Taro Watanabe and Yoshinobu Kano 60982

Closing the Verification Loop: Self-Check Captioning for Long-Paragraph Detailed Audio Captioning
Fengji Ma, Yan Rong, Xu Li, Chen Zhang, Pengfei Wan and Li Liu 60999

Copy First, Translate Later: Interpreting Translation Dynamics in Multilingual Pretraining
Felicia Körner, Maria Matveev, Florian Eichin, Gitta Kutyniok, Barbara Plank and Michael A. Hedderich 61018

CoPiT: Cognitive Pivot Translation for Digraphic Low-Resource Mongolian in the Traditional Script
Burte Bayarsaikhan, Serynn Kim and Buru Chang 61052

VCoT-SQL: Enhancing Text-to-SQL Generation through Dual Chain-of-Thought Validation Training
Rui Zhang, Yujia Su, Ruixuan Li, Yuhua Li and Yixin Su 61074

Auditing Intersectional Bias in Clinical LLMs: Sparse Autoencoders Reveal Race-HIV Stigma Circuits
Chang Lu and Yizheng Zhao 61098

PolERo: Studying Political Evasion in Romanian
Gabriel Ștefan and Sergiu Nisioi 61134

Signals Are Present, Predictions Fail: Probing Diagram Representations in Large Vision-Language Models
Haruto Yoshida, Keito Kudo, Yoichi Aoki, Ryota Tanaka, Itsumi Saito, Keisuke Sakaguchi and Kentaro Inui 61160

Agentic Rubrics: Interfacing Agents with Heterogeneous Environments using Executable Rubrics
Shunian Chen, Rui Yu, Zhengyang Tang, Yuhao Du, Zejian Xie, Songxin Zhang and Benyou Wang 61198

ResCode: Accurate Expert Compression of Sparse MoE Models via Residual-code Restoration
Hyunwoo Yang, Minjun Kim and U Kang 61227

Deriving Necessary Validity Requirements for Evaluating the Consistency of LLM Responses
Alexander Ljapunov, Niklas Kiehne and Wolf-Tilo Balke 61246

T-CLA: Taxonomy of Cultural-Linguistic Artifacts for Modeling Misalignment in NLP
Kezia Oketch, John P. Lalor, Denghui Zhang and Ahmed Abbasi 61267

Beyond Policy Optimization: A Data Curation Flywheel for Sparse-Reward Long-Horizon Planning
Yutong Wang, Ryan Ji, Kaixin Li, Baolong Bi, Tao Feng and Guillaume Adrien Sartoretti . 61285

SPLADE-COVER: Coverage-Aware Representations for Learned Sparse Retrieval
Siddharth AK Singh, Eugene Yang, Dawn Lawrie, Trevor Adriaanse and Andrew Yates . . . 61302

AdaSearch: Balancing Parametric Knowledge and Search in Large Language Models via Reinforcement Learning
Tzu-Han Lin, Wei-Lin Chen, Chen-An Li, Hung-yi Lee, Yun-Nung Chen and Yu Meng . . . 61313

TRACE: Trajectory Aware Reasoning for Multi-Turn Adversarial Conversation Evaluation.
Md Messal Monem Miah, Adrita Anika, Zhiyuan Yu and Ruihong Huang 61341

Hidden in Plain Sight: The Overlooked Significance of Canonical Elements for Extreme LLM Sparsity
Hyeondo Jang, Kwanhee Lee, Dongyeop Lee and Namhoon Lee 61372

Adopt ≠ Adapt: Longitudinal Analyses of LLM Conversations in the Wild
Rebecca M. M. Hicke and Kiran Tomlinson 61395

MENASpeechBank: A Reference Voice Bank with Persona-Conditioned Multi-Turn Conversations for SpeechLLMs
Zien Sheikh Ali, Hunzalah Hassan Bhatti, Rabindra Nath Nandi, Shammur Absar Chowdhury and Firoj Alam 61422

Reusing Latent Speech Representations for Query-Conditioned Topic Localization in Transcripts
Steffen Freisinger, Philipp Seeberger, Thomas Ranzenberger, Tobias Bocklet and Korbinian Riedhammer 61445

Persistent Teacher Anchoring for Tool-Using Agents
Hyun Bin Park, Kyungho Song, Sangmin Lee and Du-Seong Chang 61459

IGG: A Benchmark for Interactive GUI Grounding under Visibility Constraints
Kyeong Seon Kim, Jiyeon Son and Tae-Hyun Oh 61475

RUBICON: Do Language Models Know When They've Crossed the Point of No Return?
Srija Mukhopadhyay, Adnan Qidwai, Jaydeep Sen and Manish Shrivastava 61499