Findings 2026

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

Findings of the Association for Computational Linguistics: EMNLP 2026

October 24-29, 2026

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

Diamond Level

Bloomberg Engineering

Google

Platinum Level



Baidu 百度

CapitalOne

cohere

GPTZero

Megagon Labs

ORACLE

小红书 rednote

airbnb

Gold Level

蚂蚁集团 ANT GROUP

JPMorganChase

美团 美团

SAP

Tencent tether.

Silver Level

Adobe

Bronze Level

DATAmundi.ai DataoceanAI translated. Booking.com

ALICE Keenable Thomson Reuters™ oruk

Academic Ally



©2026 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

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

ISBN 979-8-89176-462-0

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

A Rising Tide Lifts All Boats: MTQE Rewards for Idioms Improve General Translation Quality
Ishika Agarwal, Zhenlin He, Dhruva Patil and Dilek Hakkani-Tür 1

GI-MIA: Group-Invariant Membership Inference Attacks Against Large Language Models
Zichen Song, Farman Ali and Tamer Abuhmed 13

GroupDPO: Memory-Efficient Group-Wise Direct Preference Optimization
Jixuan Leng, Si Si, Hsiang-Fu Yu, Vinod Raman and Inderjit S Dhillon 29

Register-based data selection for multilingual LLMs
Tomasz Galica, Amanda Myntti, Erik Henriksson, Fedor Vitiugin, Sampo Pyysalo and Veronika Laippala 52

Untangling Input Language from Reasoning Language: A Diagnostic Framework for Cross-Lingual Moral Alignment in LLMs
Nan Li, Bo Kang and Tijl De Bie 68

PRL: Prompts from Reinforcement Learning
Paweł Batorski, Adrian Kosmala and Paul Swoboda 95

Reasoning Model is Stubborn: Diagnosing and Mitigating Reasoning Rigidity in Large Language Models
Doohyuk Jang, Yoonjeon Kim, Chanjae Park, Hyun Ryu and Eunho Yang 115

Answer Bubbles: Information Exposure in AI-Mediated Search
Michelle Huang, Agam Goyal, Koustuv Saha and Eshwar Chandrasekharan 147

When Symbolic Analyses Mislead: Resolution Sensitivity in Neural Language Representations
Matthew Fried 163

ReviewScore: Misinformed Peer Review Detection with Large Language Models
Hyun Ryu, Doohyuk Jang, Hyemin S. Lee, Joonhyun Jeong, Gyeongman Kim, Donghyeon Cho, Gyouk Chu, Minyeong Hwang, Hyeongwon Jang, Changhun Kim, Haechan Kim, Jina Kim, Joowon Kim, Yoonjeon Kim, Kwanhyung Lee, Chanjae Park, Heecheol Yun, Gregor Betz and Eunho Yang 173

Two-Sided State-Space Models for Sequential Recommendation with Non-Random Multimodal Review Feedback
Ziwen Pan, Zihan Liang and Ruoxuan Xiong 209

Identity-Based Description Bias in AI-Generated Alternative Text
Jane Warren, Utsav Patel, Gary M. Weiss and Yijun Zhao 243

The Challenge of Identifying the Origin of Black-Box Large Language Models
Ziqing Yang, Yixin Wu, Yun Shen, Wei Dai, Michael Backes and Yang Zhang 259

One-shot Style Transfer LLM log-probabilities for Authorship Attribution and Verification
Pablo Miralles-González, Javier Huertas-Tato, Alejandro Martín and David Camacho 278

Ask or Assume? Uncertainty-Aware Clarification-Seeking in Coding Agents
Nicholas Edwards and Sebastian Schuster 298

Memory–Environment Dual-Grounding for Embodied Agents: Active Information Seeking Beyond Passive Observation
Jae-Woo Choi, Sehoon Oh and Ue-Hwan Kim 317

Robust Multimodal Sentiment Analysis with Incomplete Modalities via Semantic-aware Completeness based Reconstruction
Han-Jun Choi, Byunggill Joe, Saim Shin and Jin Yea Jang 339

Learning Compositional Spatio-Temporal Video Grounding with Synthetic Curriculum
Xingjian Wang, Shijian Wang, Yibo Wang, Zihao Yu, Runhao Fu, Xuelian Cheng and Zongyuan Ge 355

TSRL: Trigger-based Supervised Reinforcement Learning for Continual Agent Training
Yuqian Fu, Kaiwen Jiang, Zijun Song, Jiawei Wang, Yuanheng Zhu and Dongbin Zhao 371

SimXAI: A Simple Drop-in Layer Increases Interaction Quality and Understanding in Conversational XAI
Van Bach Nguyen, Qianli Wang, Nils Feldhus, Jörg Schlötterer and Christin Seifert 391

A Simple Method to Enhance Pre-trained Language Models with Speech Tokens for Classification
Nicolás Calbucura, José Guillen and Valentin Barriere 409

A Synthetic Request-Reply Email Corpus Annotated with Document-Level Politeness and Sentence-Level Face Acts
Roshad Alipanah, Valentin Barriere and Jorge A. Baier 419

Fact-Checking the Jailbreak: A Claim-Level Framework for Structured Misinformation Assessment
Jia Yee Tan, Rong Fu, Yongtai Liu, Kaiyi Hu and Simon James Fong 429

Cross-Domain, Multi-Task Data-to-Text Generation without In-Domain Training Data
Yifei Song, Kun Efimov-Zhang and Claire Gardent 449

Faithful yet Collusive: Why Chain-of-Thought Monitoring Cannot Detect Collusion in LLM Pricing Agents under Oligopolistic Competition
Dohun Lee and Hyunwoo Park 499

A Probing-Based Explanation of the Effect of Existential Presuppositions on Hallucinations in Large Language Models
Jonas Vaquet, Ole Pütz and Benjamin Paassen 519

Training Therapeutic Judges and Multi-Agent Systems for Human-Aligned Mental Health Support
Mizanur Rahman, Abeer Badawi, Elahe Rahimi, Laleh Seyyed-Kalantari, Frank Rudzicz, Enamul Hoque Prince and Elham Dolatabadi 536

End-to-End On-Device Quantization-Aware Training for LLMs at Inference Cost
Qitao Tan, Xiaoying Song, Ningxi Cheng, Guoming Li, Jun Liu, Lingzi Hong, Caiwen Ding, Jundong Li, Xiaoming Zhai, Shaoyi Huang, Jin Lu, Ninghao Liu, Wei Niu and Geng Yuan 556

LE-Viz: Complementing Text Serialization with Visual Encoding of Structured Temporal Data for Cross-Distribution Generalization
Ahatsham Hayat, Bilal Khan and Mohammad Rashedul Hasan 578

Egocentric Spatial Reference Resolution under ASR Noise in Traditional Chinese Conversational Navigation
Yu-Ti Huang 600

TaLK: Text-attributed Graph Dataset Distillation via Coupling Language Model with Graph-Aware Kernel
Yeongho Kim, Yeonje Choi and Kijung Shin 615

Towards Reliable Multilingual Judge Models: An Empirical Study
Irune Zubiaga, Aitor Soroa and Rodrigo Agerri 630

Souper-Model: How Simple Arithmetic Unlocks State-of-the-Art LLM Performance
Shalini Maiti, Amar Budhiraja, Bhavul Gauri, Gaurav Chaurasia, Anton Protopopov, Alexis Audran-Reiss, Michael Slater, Despoina Magka, Tatiana Shavrina, Roberta Raileanu and Yoram Bachrach 645

CorporateBench: Large-Scale Q&A Benchmarking with Temporal Knowledge Bases
Sil Hamilton, Albert Sun, Oscar J. Romero, Carl-Leander Henneking, David Mimno, Bishan Yang and Igor Labutov 667

When Vocal Tone and Literal Meaning Diverge: An Acoustic–Semantic Incongruity Study for Large Audio–Language Models
Yu-Wen Chen, William Ho, Maxim Topaz, Zoran Kostic and Julia Hirschberg 702

ORBIT: On-policy Exploration-Exploitation for Controllable Multi-Budget Reasoning
Kun Liang, Clive Bai, Xin Xu, Chenming Tang, Sanwoo Lee, Weijie Liu, Saiyong Yang and Yunfang Wu 718

ComfySearch: Autonomous Exploration and Reasoning for ComfyUI Workflows
Jinwei Su, Qizhen Lan, Zeyu Wang, Yinghui Xia, Hairu Wen, Yiqun Duan, Xi Xiao, Yang Jingsong, Lewei He and Tianyu Shi 734

iFlip: Iterative Feedback-driven Counterfactual Example
Yilong Wang, Qianli Wang and Nils Feldhus 749

Can Large Language Models Still Explain Themselves? Investigating the Impact of Quantization on Self-Explanations
Qianli Wang, Nils Feldhus, Pepa Atanasova, Fedor Splitt, Simon Ostermann, Sebastian Möller and Vera Schmitt 780

Semantic Retrieval Speculative Decoding with Intermediate Representations
Fangming Zhao, Xiaofei Yue, Fulun Ye, Yu Peng, Ziming Zhao, Lewei Jin, Wei Wu, Tianxiang Chen, Tingting Li and Jianwei Yin 810

Comparing the Framing Effect in Humans and LLMs on Naturally Occurring Texts
Gili Lior, Liron Naccache and Gabriel Stanovsky 825

M4FC: a Multimodal, Multilingual, Multicultural, Multitask Real-World Fact-Checking Dataset
Jiahui Geng, Jonathan Tonglet and Iryna Gurevych 842

Beyond Single Edits: Rethinking Knowledge Editing Evaluation with Multi-Edit Entities and Undirected Probing
Dominic Simon and Rickard Ewetz 868

CARD: Cluster-level Adaptation with Reward-guided Decoding for Personalized Text Generation
Yutong Song, Jiang Wu, Weijia Zhang, Chengze Shen, Shaofan Yuan, Weitao Lu, Jian Wang, Yu Wang, Nikil Dutt and Amir M. Rahmani 889

DySD: Dynamic Speculative Decoding for Evolving LLMs under Rapid Distributional Shifts
Xinglin Liu, Bohao Qu, Chenyu Li, Yongjie Wang, Qing Guo and Ying Wang 909

DIAL: Direct Iterative Adversarial Learning for Realistic Multi-Turn Dialogue Simulation
Ziyi Zhu, Olivier Tieleman, Caitlin A. Stamatis, Luka Smyth, Thomas D. Hull, Daniel Reid Cahn, Jinghong Chen and Matteo Malgaroli 925

Localize-Then-Decide Guarantees for LLM Judgments
Xinyu Li, Yi Zhou, Guanqun Cao, Zeyu FU, Tianjin Huang and Gaojie Jin 943

JAM: A Tiny Flow-based Song Generator with Fine-grained Controllability and Aesthetic Alignment
Liu Renhang, Chia-Yu Hung, Navonil Majumder, Dorien Herremans and Soujanya Poria 966

"Give a Positive Review Only": An Early Investigation Into In-Paper Prompt Injection Attacks and Defenses for AI Reviewers
Qin Zhou, Zhexin Zhang, Li Zhi and Limin Sun 982

ReserveGate: Reserve-Aware Evidence Scheduling for Long-Horizon Language Agents
Shuang Cao, Rui Li, Louise Ruihua Liu and Alexandre Duprey 998

TraceWeaver: Language-Conditioned Action Selection for Web Agents via Coupled Rollouts and Episodic Memory
Shuang Cao, Rui Li, Louise Ruihua Liu and Alexandre Duprey 1018

Adaptive Information Control for Search-Augmented LLM Reasoning
Siheng Xiong, Oguzhan Gungordu, Blair Johnson, James Clayton Kerce and Faramarz Fekri 1036

Enhancing Situated Cultural Reasoning in Large Language Models via Simulation Learning
Jinghao Zhang, Shisong Chen, Sihang Jiang, Yanghua Xiao, Hongwei Feng, Jiaqing Liang, Minggui HE, Shimin Tao, Mahongxia and Yilun Liu 1079

FRESCO: Benchmarking and Optimizing Re-rankers for Evolving Semantic Conflict in Retrieval-Augmented Generation
Sohyun An, Hayeon Lee, Shuibenyang Yuan, Chun-cheng Jason Chen, Cho-Jui Hsieh, Vijai Mohan and Alexander Min 1102

Reasoning Beyond Language: A Comprehensive Survey on Latent Chain-of-Thought Reasoning
Xinghao Chen, Anhao Zhao, Heming Xia, Xuan Lu, Hanlin Wang, Yanjun Chen, Wei Zhang, Jian Wang, Wenjie Li and Xiaoyu Shen 1125

CWM: Controllable White-Box Meta-Prompting for Adaptive Retrieval-Augmented Generation and Reasoning Ability
Keuntae Kim, Eunhye Jeong and Yong Suk Choi 1144

DiffPDE: Masked Diffusion Language Models as PDE Solver
Wenxuan Guo, Yuyang Hong, Lubin Fan, Zhaojin Fu, Lin Chen, Kun Ding and Shiming Xiang 1169

Prompting to Prompt: Meta-Template Learning for Transferable Prompt Optimization
Xinyu Zhang, Deyang Li, Minjun Zhao, Meng Han, Ruifeng Shi and Zhicheng Dou 1187

RADAR: Reasoning as Discrimination with Aligned Representations for LLM-based Knowledge Graph Reasoning
Bo Xue, Yuan Jin, Luoyi Fu, Jiaxin Ding and Xinbing Wang 1209

Image-Informed Post-OCR Correction with Vision-Language Models
Alan Thomas, Xianyuan Liu, Haiping Lu and Robert J. Gaizauskas 1224

ConvergeWriter: Data-Driven Bottom-Up Article Construction
Binquan Ji, Jiaqi Wang, Ruiting Li, Xingchen Han, Haibo Luo, Yiyang Qi, Shichao Wang, Yifei Lu, Yuantao Han and Feiliang Ren 1238

AdaPE: Adaptive Editing Strength for Batch Knowledge Editing in Large Language Models
Jinjin Qian, Yunqi Mi, Zeyu Song, Guoshuai Zhao and Xueming Qian 1257

Q-realign: Piggybacking Realignment on Quantization for Safe and Efficient LLM Deployment
Qitao Tan, Xiaoying Song, Ningxi Cheng, Ninghao Liu, Xiaoming Zhai, Lingzi Hong, Yanzhi Wang, Zhen Xiang and Geng Yuan 1276

On the Systematic Challenges of Culturally Loaded Machine Translation: Dream of the Red Chamber as the Cultural Lens
Yiming Wang and Jiayuan Di 1292

Towards Anticipatory Agents for Streaming Video Understanding
Haolin Yang, Feilong Tang, Lingxiao Zhao, Xinlin Zhuang, Yifan Lu, Xiang An, Ming Hu, Xiaofeng Zhang, Abdalla Swikir, Junjun He, Zongyuan Ge, Muhammad Haris Khan and Imran Razzak 1315

From Harm to Help: Turning Reasoning In-Context Demos into Assets for Reasoning LMs
Weida Liang, Zihang Fu, Haonan Wang, Zheng Nie, Yifan Zhang, Yao Tong, Tongyao Zhu, Hao Jiang, Chuang Li, Jiaying Wu and Kenji Kawaguchi 1333

Donors and Recipients: On Asymmetric Transfer Across Tasks and Languages with Parameter-Efficient Fine-Tuning
Kajetan Dymkiewicz, Ivan Vulić, Helen Yannakoudakis, Eilam Shapira, Roi Reichart and Anna Korhonen 1356

SDGBiasBench: Benchmarking and Mitigating Vision–Language Models' Biases in Sustainable Development Goals
Zihang Lin, Huaiyuan Qin, Muli Yang and Hongyuan Zhu 1374

Learning to Extract Physician-Determined Diagnosis from Clinical Narratives Using Dual Noisy Labels
Vahid Mahzoon, Elmira Talebianaraki, Solabomi Ojeniyi, Tobin Efferen, Sheetal Koul, Zivjena Vucetic, Avrum Gillespie and Slobodan Vucetic 1395

GuardrailAgent: An Agentic Framework for Jailbreak Defense
Tianze Chang, Junjie Chu, Yixin Wu, Michael Backes and Yang Zhang 1412

PACE-RAG: Patient-Aware Contextual and Evidence-Constrained RAG for Clinical Drug Recommendation
Chaeyoung Huh, Hyunmin Hwang, Jung Hwan Shin, Sungyang Jo, Jinse Park and Jong Chul Ye 1429

Looking Closer When Unsure: VIGOR, A Training-Free Framework for Mitigating Object Hallucination in Multimodal Large Language Models
Jiaxin Li, Yixiang Chen, Lefan Lai, Geng Zhao, Xiaohan Long and Xuming Hu 1463

Listening to the Echo: User-Reaction Aware Policy Optimization via Scalar-Verbal Hybrid Reinforcement Learning
Jing Ye, Xinpei Zhao, Lu Xiang, Yaping Zhang and Chengqing Zong 1485

KFS-RAG: Mitigating Database Leakage in RAG Systems with Keyword-Grounded Fact Substitution
ZiLiang Zhang, Yubo Zhu, Wei Tong, Jingyu Hua, Zijian Wang, Yuan Zhang and Sheng Zhong 1516

Reward-guided Meta-Prompt Evolving with Reflection for LLM Jailbreaking
Caiyang Yu, Wei Ju, Yun Liu, Qingqing Long, Siyu Yi, Yifan Wang, Junyu Luo and Ziyue Qiao 1533

When the Specification Emerges: Benchmarking Faithfulness Loss in Long-Horizon Coding Agents
Lu Yan, Xuan Chen and Xiangyu Zhang 1552

Decision-aware User Simulation Agent for Evaluating Conversational Recommender Systems
Yuan-Chi Li, Li-Chi Chen, Sung-Yi Wu, Yu-Che Tsai and Shou-De Lin 1572

GRIP: Granular Reward-Guided Parameter Interpolation for Efficient Reasoning
Lam So, Canhui Wu and Han Lin 1593

Do Androids Dream of Unseen Puppeteers? Probing for Conspiratorial Tendencies in Large Language Models
Francesco Corso, Francesco Pierri and Gianmarco De Francisci Morales 1606

Adaptive Guidance for Retrieval-Augmented Masked Diffusion Models
Jaemin Kim and Jong Chul Ye 1621

PSY-STEP: From Surface Problems to Underlying Beliefs through Structured Therapeutic Targets and Action Sequences
Jihyun Lee, Yejin Min, Yejin Jeon, Sungjun Yang, Hyounghun Kim and Gary Lee 1637

Benchmarking Explainable Vulnerability Detection with Calling Contexts
Jinyao Guo, Yaoyang Ye, Zhihao Guo, Mingwei Zheng, Tiantai Zhang, Xuwei Liu, Xiaolong Jin, Congyu Liu, Xiangzhe Xu, Zian Su, Ziyun Song, Zhou Xuan, Xiao Cheng, Chengpeng Wang and Xiangyu Zhang 1676

Seeing vs. Believing: Evaluating Open-Source MLLMs in Counter-Intuitive Scenes
Chen Ling, Tongwei Zhang, Hanqian Li and Nai Ding 1695

MAAM: Anchor-Preserving Compression and Contextual Calibration for Chinese Discriminatory Language Detection
YuXin Fu and Shijing Si 1713

AlignDiff: Exploiting Model-Intrinsic Information for Better Preference Data Selection
Peng Lai, He Zhu, Zhiwen Ruan, Dongdong Zhang, Yun Chen, Peng Li, Furu Wei, Yang Liu and Guanhua Chen 1734

Personalization Double Binds: When User Preferences Meet Demographic Patterns in Chatbot Behavior
Angelina Wang, Erin Beeghly, Sanmi Koyejo and Daniel E. Ho 1756

ID-Align: RoPE-Conscious Position Remapping for Dynamic High-Resolution Adaptation in Vision-Language Models
Bozhou Li and Wentao Zhang 1787

High Scores, Hidden Failures: Evaluating Codebook-Based LLM Content Analysis
Elena Senger, Yuri Campbell, Rob Van Der Goot and Barbara Plank 1805

Choosing the Right Language Mode at Inference Time for Multilingual Reliability
Ekata Mitra and Ameeta Agrawal 1821

Stable-MM-R1: Anchoring Multimodal Reasoning Dynamics via Entropy-Guided Stratification
Yimeng Ye, Shuang Chen, Wenxuan Huang, Manyuan Zhang, Kaituo Feng, Zhangquan Chen, Jiayu Chen, Yucheng Zhou, YiCheng Xiao, ZhiYuan Feng and Tianyu Shi 1840

Tri-PvP: Exposing Modality Bias in Omni-Modal Large Language Models through Perceptual-Propositional Evidence Conflicts
Yen-Ting Piao, Shu-Yun Chen, Chin-Hui Chu, Chun-Wei Chen, Shih-Yun Shan Kuan, Hung-yi Lee and Yun-Nung Chen 1854

CANVAS: Continuity-Aware Narrative Generation via Visual Agentic Storyboarding
Ishani Mondal, Yiwen Song, Mihir Parmar, Palash Goyal, Jordan Boyd-Graber, Tomas Pfister and Yale Song 1874

SeeKG: Multi-modal LLMs Understand Knowledge Graphs as Images
Lingbing Guo, Yichi Zhang, Yang Liu, Zhuo Chen and Zequn Sun 1916

Multi-MALLS: A Multilingual Benchmark for Natural Language to First-Order Logic Translation
Haodi Zhang, Xuefeng Wu, Jiawei Wen, Yichi Wang, Liang-Jie Zhang, Tao Fan and Yuanfeng Song 1933

Multilinguality as Sense Adaptation
Jan Christian Blaise Cruz, David Ifeoluwa Adelani and Alham Fikri Aji 1951

No Single Tokenizer Feature Reliably Predicts Downstream Language Model Performance
Antonio Lopardo, Luis Frentzen Salim, Shikhar Srivastava, Kevin Jiang, Mridul Sharma and Catherine Arnett 1968

Which Questions Improve Learning the Most? Utility Estimation of Questions with LM-based Simulations
Dong-Ho Lee, Hyundong Justin Cho, Jonathan May and Jay Pujara 2015

LogicSkills: A Structured Benchmark for Formal Reasoning in Large Language Models
Brian Rabern, Philipp Mondorf and Barbara Plank 2035

CultureTalk: Towards Multi-turn Cultural Competence in Large Language Models
Minzhi Li, Jiayu Han, Oleg Poliannikov, Vipul Agarwal, Dezhi Hong and Mingxian Wang . 2048

VideoMem: Enhancing Ultra-Long Video Understanding via Adaptive Memory Management
Hongbo Jin, Jingqi Tian, Jiayu Ding, Qiaoman Zhang, Wenhao Zhang, Siyi Xie, Yang Liu and Sijie Cheng 2066

End-to-End Argument Mining with LLMs: Representation Insights into Multitask Learning
Zhiqiang Shi and Oana Cocarascu 2082

Topological Ordering for Explicit Global Structural Reasoning in Argument Relation Identification
Zhiqiang Shi and Oana Cocarascu 2099

HiCo-EL: Encoding Ontology Structure into Bi-Encoder Entity Linking
Albert Lee, Prasanth Bathala, Puru Malhotra, Christophe Ye and Cassie S. Mitchell 2113

Cross-lingual Clustering of News Stories using Embeddings-based Community Detection
Nikolaos Nikolaidis, Kristina Kovacikova and Ion Androutsopoulos 2124

BEACON: Bayesian Optimal Stopping for Efficient LLM Sampling
Guangya Wan, Zixin Stephen Xu, Sasa Zorc, Manel Baucells, Mengxuan Hu, Hao Wang and Sheng Li 2147

GONE: Structural Knowledge Unlearning via Neighborhood-Expanded Distribution Shaping
Chahana Dahal, Ashutosh Balasubramaniam and Zuobin Xiong 2173

MultivationBench: A Benchmark for Multimodal Sequential Motivation Reasoning
Kawai Chung, Chunkit Chan, Yauwai Yim, Yuxuan Liu, Haochen Shi, Weiqi Wang, Qing Zong, Tianshi Zheng, FU Yixuan, Wong Kai Chung, Hao Liang, Yifan Gao, Xi Yang, Janet Hui-wen Hsiao and Yangqiu Song 2194

Multi-Expert Routing for Event Graph Completion
Qing Zhang, Wei Cai, Chao Liang and Bang Wang 2229

SHIFT: Semantic Harmonization via Index-side Feature Transformation for Multilingual Information Retrieval
Youngjoon Jang, Seongtae Hong, Hyeonseok Moon and Heuiseok Lim 2247

Uncertainty-o: One Model-agnostic Framework for Unveiling Uncertainty in Large Multimodal Models
Ruiyang Zhang, Junhao Zhang, Hao Fei and Zhedong Zheng 2270

A Retraining-Free Framework for Analyzing Training Order Effects in Large Language Models
Hao Yang, Haoxuan Li, Mengyue Yang, Xu Chen and Mingming Gong 2286

MuCoCo: Automated Consistency Testing of Code LLMs
Chua Jin Chou, Khant That Lwin and Ezekiel Soremekun 2307

Fortifying Hallucination Detection Under Distribution Shift in Short-Form QA
Ti Tze Hong, Saffron Salmah Yen Lim and Maarten Sap 2327

EmoRL: Enhancing Emotional Fidelity of Role-Playing Agents via Reinforcement Learning
Qiming Feng, Xiaolong Wang, Qianqing Tang, Qingqiu Li, Yuejie Zhang, Rui Feng, Tao Zhang and Shang Gao 2351

A Comprehensive Analysis for Visual Object Hallucination in Large Vision-Language Models
Liqiang Jing, Hardy Chen, Ehsan Aghazadeh, Xin Eric Wang and Xinya Du 2365

Can Video Large Multimodal Models Think Like Doubters—or Double-Down: A Study on Defeasible Video Entailment
Yue Zhang, Jilei Sun, Yunhui Guo and Vibhav Giridhar Gogate 2379

HiMed: Incentivizing Hindi Reasoning in Medical LLMs
Dingfeng Jiang, Han Yan, Chenze Ma, Amit Kumar Jaiswal, Ang Li, Yunxiang Jiang, Xinlei Xiong, Juhao Liang, Hongru Xiao, Xiang Li, Fan Bu, Jiale Han, Ruchir Gupta, Prayag Tiwari and Benyou Wang 2399

Demystifying Reinforcement Learning for Long-Horizon Tool-Using Agents: A Comprehensive Recipe
Xixi Wu, Qianguo Sun, Ruiyang Zhang, Chao Song, Junlong Wu, Yiyan Qi and Hong Cheng 2429

Cascaded Batch Prompting
Sho Hoshino and Peinan Zhang 2462

LongReasonArena: A Long Reasoning Benchmark for Large Language Models
Jiayu Ding, Shuming Ma, Lei Cui, Nanning Zheng and Furu Wei 2470

Equitable Alignment of LLMs Across Languages with Regularized Direct Preference Optimization
Thibaud Leteno, Antoine Gourru, Charlotte Laclau and Christophe Gravier 2486

Adaptive Visual Token Pruning for Efficient Multi-Image Understanding in Large Visual Language Models
Rongyang Zhang, Chengqiang Lu, Cong Li, Hongchao Gu, Tingjia Shen, Xuyang Zhi, Qimeng Wang, Yan Gao, Yi Wu, Yao Hu, Hao Wang and Enhong Chen 2496

The Reward Model Selection Crisis in Personalized Alignment
Fady Rezk, Yuangang Pan, Chuan-Sheng Foo, Xun Xu, Nancy F. Chen, Henry Gouk and Timothy Hospedales 2510

Reader Proficiency Shapes Layer-wise Surprisal Profiles
Akio Hayakawa and Horacio Saggion 2529

VDAHM: Vision-Delimiter-aware Attention Head Modulation for Enhancing Multi-Image Understanding in Large Vision-Language Models
Zhibiao Liang, Junyu Cheng and Ziqian Zeng 2542

SafeMo: Linguistically Grounded Unlearning for Trustworthy Text-to-Motion Generation
Yiling Wang, Zeyu Zhang, Yiran Wang and Hao Tang 2562

FedCoT: Communication-Efficient Federated Reasoning Enhancement for Large Language Models
Chuan Li, QianYi Zhao, Fengran Mo and Cen Chen 2586

PILAR: A Page-Grounded Unified Evidence Representation via an Entity-Linked Assertion Graph for Open-Domain QA Agents over Multimodal Document Corpora
Joongmin Shin, Gyuho Shim, Jung-Hun Lee and Jaehyung Seo 2606

AgentPack: A Dataset of Code Changes, Co-Authored by Agents and Humans
Yangtian Zi, Zixuan Wu, Aleksander Boruch-Gruszecki, Jonathan Bell and Arjun Guha 2631

MineCEraft: Evaluating Language Models as Construction Engineers in the World of Minecraft
Sewoong Lee, Risham Sidhu, Julia Hockenmaier and Yoonhwa Jung 2643

OPEN-SWE-TRACES: Advancing Dual-Mode Multilingual Distillation for Software Engineering Agents
Wasi Uddin Ahmad, Nikolai Ludwig, Somshubra Majumdar and Boris Ginsburg 2668

Spatial Audio Question Answering and Reasoning on Dynamic Source Movements
Arvind Krishna Sridhar, Yinyi Guo and Erik Visser 2680

PTCG: Persona-guided Tree-based Counterargument Generation
Eunbeen Son, Yohan Jo, Joonsuk Park and JinYeong Bak 2705

Tree-of-Concerns: Hierarchical Multi-Agent Debate for Unstated-Limitation Extraction in Scientific Critique
Sahil Mishra, Niranjan Rajeev and Tanmoy Chakraborty 2735

A Survey of Quantized Graph Representation Learning: Connecting Graph Structures with Large Language Models
Qika Lin, Zhen Peng, Kaize Shi, Kai He, Yiming Xu, Jian Zhang, Erik Cambria and Mengling Feng 2765

Ability-Aligned Skill Evolution for LLM Agents via Reinforcement Learning
Fei Huang, Tianyu Chen, Zhongyu He, Yuanfan Li, Meng Hsuan Yu, Xingyang Li, Lu Pan, Ke Zeng and Xunliang Cai 2779

DRNash: Efficient and Robust Multi-Task Learning for Financial Sentiment Analysis
Hongcheng Ding, Xuanhuang Liu, Deshinta Arrova Dewi, Shamsul Nahar Abdullah, Qingyu Liu and Xuanze Zhao 2795

Robustness as an Emergent Property of Task Performance
Shir Ashury-Tahan, Ariel Gera, Elron Bandel, Michal Shmueli-Scheuer and Leshem Choshen 2847

Mechanism-Level Evaluation for Vision-Language Models: Controlled Activation-Replacement Diagnosis of Gender Bias
Zhipeng Zhao, Wenxu Wang, Peishun Liu and Ruichun Tang 2870

What Do Hallucinations Reveal About Multimodal Reasoning? Diagnosing Visual Grounding Failures via Contrastive Decoding Probes
Zhipeng Zhao, Wenxu Wang, Peishun Liu and Ruichun Tang 2889

Utilizing Geometric Structural Features for Functional RNA Design
Jiyue Jiang, Yuhan Chen, Xiangyu Shi, Yihan Ye, Han Huang, Yanyu Chen, Yuheng Shan, Dongchen He, Jiao Yuan and Yu Li 2913

Better with Experience: Self-Evolving LLM Agents for Evidence-Grounded Health Community Notes
Zihang Fu, Fanxiao Li, Jianyang Gu, Haonan Wang, Preslav Nakov, Bryan Hooi, Min-Yen Kan and Jiaying Wu 2924

VRR: Verification-Guided Retrieval Revision for Multi-Step QA
Zhijie Li, Lei Wang, Tianyuan Li, Bangju Han, Ahtamjan Ahmat, Rui Dong and Zhen Wang2962

OCR-MetaReasoning Benchmark: Evaluating the Meta-Reasoning Ability of MLLMs in Text-Rich Image Understanding
Gengxu Li, Yuan Wu and Yi Chang 2993

VCIFBench: Evaluating Complex Instruction Following for Video Understanding
Huangchen Xu, Yuan Wu and Yi Chang 3034

From Isolation to Alignment: Unified LoRA for Efficient Multi-Task Learning
Jinda Liu, Yi Chang and Yuan Wu 3065

ForeDreamer: A Self-Evolving Dual-Agent Memory Architecture for Future Event Prediction
Linhao Zhong, Zongze Du, Linyu Wu, Yu Bo, Hourong Li, Chenchen Jing, Hao Chen, Yuling Xi and Chunhua Shen 3085

EduDial: Constructing a Large-scale Multi-turn Teacher–Student Dialogue Corpus
Shouang Wei, Min Zhang, Xin Lin, Bo Jiang, Zhongxiang Dai and Kun Kuang 3113

SkillForge: Compositional Skill Synthesis with Verification-in-the-Loop for Generating Formally Verified Dafny Programs
Yanming Liu, Xinyue Peng, Jiannan Cao, Xinyi Wang and Jinbo Su 3135

From Layers to Submodules: Rethinking Granularity in Replacement-Based LLM Compression
Elia Cunegatti, Marcus Vukojevic, Erik Nielsen and Giovanni Iacca 3154

Macro: Enhancing Multilingual Counterfactual Explanations through Alignment-as-Preference Optimization
Yilong Wang, Qianli Wang, Bohao Chu, Yihong Liu, Jing Yang and Simon Ostermann 3176

Beyond Correctness: Benchmarking LLMs for Algorithmic Program Optimization
Yehua Lin and Yin Chen 3211

HINT-SD: Targeted Hindsight Self-Distillation for Long-Horizon Agents
Woongyeong Yeo, Yumin Choi, Taekyung Ki and Sung Ju Hwang 3227

EventKD: Event-Aware Knowledge Distillation for Generative Event Extraction
Hoang Thanh Dat, Quoc Phong Dao, Nam Le Hai, Franck Dernoncourt, Linh Ngo Van, Trung Le and Thien Huu Nguyen 3240

Detecting Hallucinations for Large Language Model-Based Knowledge Graph Reasoning
Xinyan Zhu, Yaoqi Liu, Yue Gao, Huadong Ma, Cheng Yang and Chuan Shi 3254

ProgRouter: Online Progress-Guided Orchestration for Multi-Agent LLM Workflows under Quality-Cost Tradeoffs
Songyuan Li, Ahmed M. Abdelmoniem and Shiqiang Wang 3272

LintEval: Benchmarking LLMs on Linter-Guided File Refactoring
Tongwei Deng, Jinyu Zhu, Yue Cao, Chenxu Wang, Yanzhen Zou and Bing Xie 3288

Complementary Lattice Correction for Post-Training Quantization
Tung Nguyen, Quang Hung Tran, Linh Ngo Van, Dinh Thai Hoang, Diep Nguyen, Trung Le and Duc Anh Nguyen 3309

xVerify: Efficient Answer Verifier for Reasoning Model Evaluations
Ding Chen, Qingchen Yu, Pengyuan Wang, Mengting Hu, Wentao Zhang, Zhengren Wang, Bo Tang, Feiyu Xiong, Xinchi Li, Chao Wang, Mingchuan Yang and Zhiyu li 3335

Hiding in Plain Sight: A Steganographic Approach to Stealthy LLM Jailbreaks
Jianing Geng, Biao Yi, Zekun Fei, He Ruiqi, Zhihao Xu, Juncheng Hu, Tong Li, Zheli Liu and Lihai Nie 3372

Dynamic Alignment Compensation for Hallucination Mitigation in Large Vision-Language Models
Kairong Yu, Zixin Zhu, Le Yu and Hongwei Wang 3392

Unexplored flaws in multiple-choice VQA make benchmarking unreliable
Fabio Rosenthal, Sebastian Schmidt, Dr. Thorsten Graf, Thorsten Bagdonat, Stephan Günnemann and Leo Schwinn 3412

DICS: Exploring Data Intrinsic Consistency for Visual Instruction Selection
Yuyang Hong, Jinhui Guo, Jiaqi Gu, Lubin Fan, Ruixiang Wang, Kun Ding, Yue Wu, Shiming Xiang and Jieping Ye 3452

The Impostors Method as a Statistical Test for Authorship Verification With Large Language Models as Sampling Source for Hard Negatives
Klara M. Gutekunst, Yaron Winter and Martin Potthast 3473

Text-Anchored Semantic Perturbations for Transferable Jailbreak Attacks on Multimodal Large Language Models
Wenyun Li, Guiping Cao, Xiangyuan Lan and Zheng Zhang 3500

STEER: Structured Text Editing via Edit-Chain Reinforcement Learning
Yucheng Lin and Jiaojiao Jiang 3516

SIGHT: Gaze-Guided Attention-Head Selection for Head-Sparse PEFT
Xinyu Gao, Shaonan Wang, Jiajie Zou, Li Zhao and Nai Ding 3529

Reasoning about In-Context Samples for Machine-Translation
Maxime Bouthors, Josep Crego and François Yvon 3546

Is Human Annotation Necessary? Iterative MBR Distillation for Error Span Detection in Machine Translation
Boxuan Lyu, Haiyue Song and Zhi Qu 3567

SA-Bench: Evaluating Semantic Alignment in LLM-Based Paper Reproduction
Xue Hu, Pan Zewei, Zeli Su, Zhou Liu and Wentao Zhang 3578

Knowing What to Solve Before How: Preplan Empowered LLM Mathematical Reasoning
Shaojie Wang and Liang Zhang 3608

Corpus2Skill: Distilling Enterprise Knowledge into Navigable Agent Skills for QA and RAG
Yiqun Sun, Pengfei Wei and Lawrence B. Hsieh 3624

LazyMuon: Exploiting Angular Stability to Reduce Newton-Schulz Overhead in Muon for Efficient Large Model Training
Kainan Lin, Rongwei Lu, Haotian Dong, Le Chen, Jingyan Jiang, Huan Yang and Zhi Wang 3646

BiNSGPS: Geometry Problem Solving via Bidirectional Neuro-Symbolic Interaction
Qi Wang, Peijie Wang, Fei Yin and Cheng-Lin Liu 3660

ReproAgent: Contract-Guided Paper-to-Code Reproduction
Xue Hu, Pan Zewei, Zhongyuan Wang, Zhou Liu, Zeli Su and Wentao Zhang 3681

How does Multilingual Finetuning of Instruction-Tuned Large Language Models Affect Languages in Layer-Wise Hidden Representations?
Hyemi Jo and Hyopil Shin 3703

Code Monitor Red Teaming for Public-Test-Passing Code
Junchi Liao, Jiawen Deng and Fuji Ren 3720

MMPCBench: Benchmarking Multimodal Large Language Models on Proactive Critique of Flawed Inputs
Jinzhe Li, Gengxu Li, Jinnan Li, Yuan Wu and Yi Chang 3741

Quantifying Error Tolerance in Synthetic Data: An Atomic-level Operand vs. Operator Perturbation Study
Jiaxiang Liu, Chenhao Yuan, Shuwen Xu, Boxuan Xing, Xiusheng Huang, Yinhao Xu, Hao Liu, Wenhao Teng, Xiangwen Liao, Pengfei Cao, Jun Zhao and Kang Liu 3775

A Survey on Knowledge Distillation of Vision-Language Models
Hoang Tran Vuong, Dat Dao, Pham Khanh Chi, Minh-Phuc Truong, Quyen Tran, Linh Ngo Van, Thien Huu Nguyen, Thanh-Toan Do and Trung Le 3806

LaRe: Latent Refocusing for Multimodal Reasoning
Jizheng Ma, Xiaofei Zhou, Geyuan Zhang, Yanlong Song and Han Yan 3847

ReTrace: Visual Backtracking for Efficient Multimodal Large Language Models
Guangzong Si, Dong Wang, Yifan Yu, Miao Liu, Panwang Pan and Wentao Zhu 3869

Where Does the Sound Go? Tracing Acoustic Information Loss in Audio-Conditioned LLMs
Song-ha Jo, Sehyun Lee, Soyoon Kim, Jaesik Choi and Sanghyuk Choi 3898

URGen: Uncertainty-guided Retrieval-augmented Generation for Guideline-Grounded Medical Question Answering
Mengieong Hoi, Chen Zhao, Chaoyu Lei, Huifang Zhou and Zhedong Zheng 3918

SemOPT: Fixing Semantic Errors in LLM-based Optimization Modeling via Reward-Guided Search
Zetong Zhou, Wentao Zhang, Jingyuan Wang, Yifan Yang, Zizhuo Wang and Shixi Hu 3937

TISR: Breaking Vision–Language Models via Text–Image Segmentation and Reassembly
Qi Lu, Simeng Qin, Xiaoshuang Jia, Jie Liao, Yihao Huang, Tianhang Zheng, Ke Ma and Xiaojun Jia 3954

Generating Constructive Feedback on Stories via Reinforcement Learning
Maja Stahl, Timon Ziegenbein and Henning Wachsmuth 3976

Beyond Over-Refusal: Defending Indirect Prompt Injection via Latent Instruction Manifolds
Jiahao Chen, Rui Yin, Xinfeng Li, Qianli Ma, Tianyu Du, Zhihui Fu, Jun Wang, Zhaoxiang Wang and Shouling Ji 3995

Reasoning or Rambling? Exploring the Effect of Thinking on Agent Persuasion
Haodong Zhao, Jidong Li, Zhaomin Wu, Tianjie Ju, Zhuosheng Zhang, Bingsheng He and Gongshen Liu 4002

ChemEval2: A Diagnostic Benchmark for Evaluating Tool-Integrated Reasoning in Chemistry
Yuqing Huang, Chuang Tang, Huadong Liang, Xianghui Fan, Hao Wang, Yin Xu, Yang Xia, Weiyi Jiang, De-Guang Liu, Feiyang Xu, Xin Li, Shijin Wang, Mingjun Xiao, Defu Lian, Guiquan Liu and Enhong Chen 4026

Extending Fill-In-the-Middle with Instructions for Steerable Code Completion
Zhensu Sun, Chengran Yang, Chao Peng, Pengfei Gao, Xiaoning Du, Li Li and David Lo . . 4050

Sustaining Cross-Modal Dependency for Generalizable Vision-Language Model Tuning
Chenhao Ding, Jizhou Han, SongLin Dong, Xinyuan Gao, Qiang Wang, Lixin Xiu, Wei Ke, Yihong Gong and Hideki Nakayama 4067

On the Challenges and Opportunities of Learned Sparse Retrieval for Code
Simon Lupart, Maxime Louis, Thibault Formal, Hervé Déjean and Stéphane Clinchant 4096

Membership Inference in Fine-tuned Diffusion Language Models via Token-level Memorization Asymmetry
Shengfang Zhai, Leo Marchyok, Yuling Shi, Huanran Chen, Yinpeng Dong, Jiaheng Zhang and Sanghyun Hong 4113

HarmEvoSkill: Skill-Memory Self-Evolution for Zero-shot Multimodal Harmful Meme Detection
Jianyu Zhou and Weihai Lu 4131

MolQuest: A Benchmark for Agentic Evaluation of Abductive Reasoning in Chemical Structure Elucidation
Taolin Han, Shuang Wu, Jinghang Wang, Zhou Yuhao, Renquan Lv, Bing Zhao, HU Wei and Lin Qu 4147

From Evasive Servants to Dignified Peers via Dual Preference Optimization
Xiangqi Wang, Yue Huang, Haomin Zhuang, Kehan Guo and Xiangliang Zhang 4164

Why Are LLM Backdoor Defenses Fragmented? A Feature-Level Explanation with Sparse Autoencoders
Yizhe Zeng, Chenxu Niu, Wei Zhang, Hao Huang, Yunpeng Li, Handongxu, Dan Du, Cheng Hong, Hequn Xian and Yuling Liu 4187

RoboSemanticBench: Diagnosing Semantic Grounding in Action Prediction for VLA Models
Bin Yu, Yao Zhang, Haishan Liu, Shijie Lian, Yuliang Wei, Xiaopeng Lin, Zhaolong Shen, Changti Wu, Ruina Hu, Bailing Wang, Cong Huang and Kai Chen 4209

When to Adapt: Conditional Memory Adapters for Retention-Preserving Domain Specialization
Jiayu Hou and Lei Wang 4224

SQD: Sub-Query Decomposition for Reliable Jailbreak Evaluation in Large Language Models
Junjie Yang, Fenghua Weng, Yue Xu and Wenjie Wang 4246

Bridging Reasoning and Retrieval: A Large-Small Model Collaboration Framework for Multimodal Entity Alignment
Xiongnan Jin, Jianming Lin, Tiantian Mei, Tiantian Zhu, Changxin Huang, Jinpeng Chen, Chaoqun Hong and Jianqiang Li 4262

Skill-3D: Evolving Scene-Aware Skills for Agentic 3D Spatial Reasoning
Haoyuan Li, Zhengdong Hu, Jun Wang, Hehe Fan and Yi Yang 4273

Attend to Evidence: Evidence-Anchored Spatial Attention Supervision for Multimodal RLVR
Ruina Hu, Chen Wang, Lai Wei, Bai Jionghao, Bin Yu, Weiran Huang, Kai Wang and Yue Wang 4295

Long-Short Memory Steering to Mitigate Unfaithful Self-Brainwashing
Mingjie Li, Michael Backes, Yang Zhang and Yisen Wang 4314

SynapseHSG: One-Shot Hyper-Semantic Grounding for Long-Context Multimodal Question Answering
Sherricane Jiayu Xu, Zhijie Shen, Yongqiang Liu, Xiaojun Jing and Shuangyin Li 4326

FrameSkip: Learning from Fewer but More Informative Frames in VLA Training
Bin Yu, Shijie Lian, Xiaopeng Lin, Zhaolong Shen, Yuliang Wei, Changti Wu, Hang Yuan, Haishan Liu, Bailing Wang, Cong Huang and Kai Chen 4345

Role-Agent: Bootstrapping LLM Agents via Dual-Role Evolution
Xucong Wang, Ziyu Ma, Shidong Yang, Tongwen Huang, Pengkun Wang, Yong Wang and Xiangxiang Chu 4360

StepOPSD: Step-Aware Online Preference Self-Distillation for Agent Reinforcement Learning
Yanfei Zhang, Xu Lin and Chenglin Wu 4380

Refine: A General Context-Aware Agent for Evolving Program Repair Agents
Anvith Pabba, Simin Chen, Alex Mathai, Anindya Chakraborty and Baishakhi Ray 4391

Vulnerable Code Search: Transferable Attack for Code Language Models
Kaicheng Wang, Liyan Huang, Jesse Thomason and Weihang Wang 4410

PatAlign: Bridging the Gap between Scientific Papers and Patent Descriptions via Iterative Preference Optimization
Zheming Fan, Xuemeng Song, Min Yang and Linyi Yang 4434

ICICLE: Expanding Retrieval with In-Context Documents
Yu-Chen Den, Yung-Yu Shih, Zhi Rui Tam, Kuan-Yu Chen, Pu-Jen Cheng, Yun-Nung Chen and Eugene Yang 4449

TrajSelector: Harnessing Latent Representations for Effective and Robust Best-of-N Selection in Large Reasoning Models
Bin Yu, Xinming Wang, Shijie Lian, Haotian Li, Changti Wu, Ruina Hu, Bailing Wang, Yuliang Wei and Kai Chen 4467

SKILL0: In-Context Agentic Reinforcement Learning for Skill Internalization
Zhengxi Lu, Zhiyuan Yao, Jinyang Wu, Chengcheng Han, Xierui Song, Xunliang Cai, Weiming Lu, Jun Xiao, Yueting Zhuang and Yongliang Shen 4485

CAGE: Concept Algebra for Guardrail Evasion and Jailbreak Attacks via Lexical-Semantic Decoupling
Zekai He, Yuxiang Chen, Jun Shao, Xiao Song and Haibo Hong 4506

TwinBrainVLA: Unleashing the Potential of Generalist VLMs for Embodied Tasks via Asymmetric Mixture-of-Transformers
Bin Yu, Shijie Lian, Xiaopeng Lin, Zhaolong Shen, Changti Wu, Yuzhuo Miao, Xinming Wang, Bailing Wang, Cong Huang and Kai Chen 4533

SkillLadder: Multi-User Skill Evolution with Compliance-Oriented Synchronization and Quality-Gated Synthesis
Nan Chen, Yi Wang, Bin Benjamin Zhu and Haoyi Xiong 4553

VIG: Visual Information Gain as a Reward Signal for Multimodal Chain-of-Thought Compression
Wen Luo, Xiaohan Yi, Xiaotao Huang and Liqun Huang 4574

Toward Native Multimodal Modeling: A Roadmap
Siyu An, Junru Lu, Junnan Dong, Qiufeng Wang, Yinghui Li, Weizhi Fei, Zichao Yu, Zheng Yuan, Biao Liu, Haopeng Wang, Renzhao Liang, Yixuan Yang, Yunhang Shen, Bo Ke, Keyu Chen, Linhao Luo, Difan Zou, Xiao Huang, di Yin, Ruizhi Qiao and Xing Sun 4594

LLaDA-o: An Effective and Length-Adaptive Omni Diffusion Model
Zebin You, Xiaolu Zhang, Jun Zhou, Chongxuan Li and Ji-Rong Wen 4645

Skill or Skip? Learning Selective Skill Invocation in Agentic Tasks via Dual-Granularity Preference Learning
Chishui Chen, Jiaye Lin, Te Sun, Yi Yang, Junxi Wang, Cong Qin, Yangen Hu, Lu Pan and Ke Zeng 4665

Group Structural Proximity Shapes Exploration-Exploitation in Tool-Use Reinforcement Learning
Lexiao Zou, Yitong Yao, Gongwei Chen, Wenmin Deng, Chao Wang, Shuangyong Song and Miao Zhang 4682

Cleansing the Artificial Mind: A Self-Reflective Detoxification Framework for Large Language Models
Kaituo Zhang, Zhimeng Jiang and Na Zou 4699

Causal Intervention with LMMs: Generative Text Facilitates Fine-Grained Pedestrian Anomaly Behavior Retrieval
Weifeng Xu, Shaofei Huang, Yaxiong Wang and Zhedong Zheng 4723

Don't Generate, Just Score: Enhancing Topic Coherence with Masked Semantic Alignment
Tien-Phat Nguyen, Minh Chu Xuan, Phung Gia Huy, Linh Ngo Van, Duc Anh Nguyen, Trung Le and Thien Huu Nguyen 4739

Aligning Progress and Feasibility: A Neuro-Symbolic Dual Memory Framework for Long-Horizon LLM Agents
Bin Wen, Ruoxuan Zhang, Yang Chen, Hongxia Xie and Lan-Zhe Guo 4767

STAGE-RAG: Continuous Entity State Evolution over Dual-Space Subgraphs for Multi-Hop Reasoning
Xinran Xu, Shaowen Chen, Yu Ma, Fulin Lin, Zhaofen Wu and Hongwei Wang 4792

Language-Guided Time-Series Event Detection with Zero-Shot Cross-Modal Grounding
Tianjun Hou, Xiqing Chang and Sky Chenwei Wan 4813

Translation Asymmetry in LLMs as a Data Augmentation Factor: A Case Study for 6 Romansh Language Varieties
Jannis Vamvas, Ignacio Pérez Prat, Angela Heldstab, Dominic P. Fischer, Sina Ahmadi and Rico Sennrich 4837

LitSeg: Narrative-Aware Document Segmentation for Literary RAG
Ruikang Zhang, Zhanni Chen, Yiqiao Cai and Qi Su 4862

Not Truly Multilingual: Script Consistency as a Missing Dimension in VLM Evaluation
Prabhjot Singh, Bhushan Pawar, Madhu Reddiboina and Rajvee Sheth 4893

Entropy-KL Divergence-based Token Masking: A Novel Approach for Selective Fine-tuning of Large Language Models
Qi Liu, Mingdi Sun, Yongyi He, Zhi Zheng, Tong Xu, Yi Zheng, Zhefeng Wang and Enhong Chen 4910

AgenticRec: A Recommendation-Oriented Agentic Framework with Progressive Tool-Integrated Reasoning Optimization
Tianyi Li, Zixuan Wang, Guidong Lei, Xiaodong Li and Hui Li 4929

HardMTBench: Stress-Testing Chinese-English Translation on Knowledge-Intensive Domains
Zheng Li, Mao Zheng, Mingyang Song and Tianxiang Fei 4947

TTS-Guard: Black-Box Ownership Verification of Text-to-Speech Models via Adaptive Adversarial Speaker-Pair Fingerprints
Xubin Yue, Zhenhua Xu, Zhebo Wang, Mengting Li, Zijie Zhou, Wenpeng Xing, Dezhang Kong and Meng Han 4958

Trinity: Human–Agent–Environment Alignment for Embodied Task Planning
Hyungmin Kim, Hobeom Jeon, Dohyung Kim, Jinhyeok Jang and Minsu Jang 4974

Asymmetric On-Policy Distillation: Bridging Exploitation and Imitation at the Token Level
Nan Jia, Haojin Yang, Xing Ma, Jiesong Lian, Shuailiang Zhang, Weipeng Zhang, Ke Zeng, Xunliang Cai and Zequn Sun 5014

BID-SPO: Rethinking Multi-Agent Prompt Optimization as Information Contracting via Bidding-based Shapley Values
Weiqiang Jin, Yuxin Cheng, Tingting Sun, Shixiang Tang, Jing Gao, Jinhu Qi, Wentao Zhang, Junli Wang, Biao Zhao, Guoshuai Zhao and Guang Yang 5029

M3-TTS: Native Two-Stream Multi-Modal Diffusion Transformer for Zero-Shot Speech Synthesis
Xiaopeng Wang, Chunyu Qiang, Ruibo Fu, Zhengqi Wen, Yuxin Guo, Xuefei Liu, Yukun Liu, Yuzhe Liang, Kang Yin, Teng Ma, Yushen Chen, Yuankun Xie, Chenxing Li and Changsheng Li . 5061

A Survey of Secure Retrieval-Augmented Generation
Yuming XU, Mingtao Zhang, Zhuohan Ge, Haoyang LI, Nicole HU, Yongqi Zhang, Zhiyuan Wen, Chen Jason Zhang, Li Qing and Lei Chen 5072

First Things First: Teaching LLM-Based Agents to Prioritize Must-Haves before Nice-to-Haves
Tianjie Ju, Xinyue Xu, Wanxuan Sun, Lingxiao Diao, Gongshen Liu, Zhuosheng Zhang and Cheng Yang 5096

Beyond Consensus: Downward Bias and Role Asymmetry in Multi-Agent LLM Judges for Subjective Evaluation
Minsoo Song, Chanwoo Kim, Sugyeong Eo and Chanjun Park 5116

Skill Following: Evaluating Actual Skill Use in Retrieval-Enabled LLM Agents
Seonghyeon Cho and Chanjun Park 5134

Rethinking Standard Compliance in the LLM Era
Joseph Marvin Imperial, Matthew D Jones and Harish Tayyar Madabushi 5156

MOSAIC: A Large-Scale Multimodal Open-Course Segmentation and Alignment Corpus in Chinese
Yuming Ai, Shuai Fan, Hua Xu and Fang Kong 5176

COPE: Continual Personalization of LLMs under Sparse User Feedback via User Embeddings and Self-Evaluation
Ruike Cao, Fugen Yao, Liang Dong, Jian Xu, Guanjun Jiang and Li Xiao 5196

Coder as Editor: Code-driven Interpretable Molecular Editing
Wenyu Zhu, Chengzhu Li, Xiaohe Tian, Yifan Wang, Yinjun Jia, Jianhui Wang, Bowen Gao, Haichuan Tan, Ya-Qin Zhang, Wei-Ying Ma and Yanyan Lan 5216

SD-ICL: Stage-Decomposed In-Context Learning for Tool Use in LLM Agents
Kangci Xiong, Xiaohan Ge, Yangyang Qin and Fusheng Jin 5237

Speech-XL: Towards Long-Form Speech Understanding in Large Speech Language Models
Haoqin Sun, Chenyang Lyu, Shiwan Zhao, Xuanfan Ni, Xiangyu Kong, Longyue Wang, Weihua Luo and Yong Qin 5255

Per-Matrix Optimality Is Not Enough: Three-Level Optimization for Low-Rank LLM Compression
Huicheng Zhang, Xiyao Feng, Ze-Tong Li, Chengkai Zhu, Xiao Shi, Xiwei Pan, JinGuo Liu, Ge Bai and Xin Wang 5268

Super Library Agent: Joint Generation and Maintenance of Multiple Applications Beyond the Single Codebase
Daegyu Sung, Yukyeong Lee, Geon Park, Yumin Choi and Sung Ju Hwang 5284

TRACE: Single-Pass Decoding-Trace Risk Localization for Generation Calibration
Yuebin Xu, Xuemei Peng, Junlan Chen, Zhiyi Chen and Zeyi Wen 5318

SEAM: Spectral Embedding Alignment and Merging for Low-Resource TTS
Walaa Othman, Baher Mohammad, Stamatios Lefkimmiatis and Ammar Ali 5336

Global Evolutionary Steering: Refining Activation Steering Control via Cross-Layer Consistency
Xinyan Jiang, Wenjing Yu, Di Wang and Lijie Hu 5348

FastSLM: Hierarchical Temporal Abstraction for Efficient Long-Form Speech Adaptation
Junseok Lee and Chang-Jae Chun 5374

CLARA: Can AI Assess Developmental Appropriateness in Children's Stories?
Sijing Yin, Zirui Wang, Qian Liu and Jiamou Liu 5390

Statistical Priors for Implicit Preferences: Decoupling Skill Selection as a Local Harness in Personal Agents
Zeyu Gan, Huayi Tang and Yong Liu 5404

Beyond Isolated Behaviors: Hierarchical User Modeling for LLM Personalization
Liang Wang, Xinyi Mou, Xiaoyou Liu, Tiannan Wang, Yuqing Wang and Zhongyu Wei 5432

TAKE: Task-Aware Chunked KV Cache Eviction for Efficient Long-Context LLM Prefill
Rui Wang, Nan Jia, Jiahui Li, Qingyi Yang, Yixue Hao and Long Hu 5449

TeacherGRPO: Closing the Capacity Gap in Reasoning Distillation via Teacher Alignment
Zhenyu Lei, Zihan Chen, Yaochen Zhu, Shangbin Feng, Zaiyi Zheng, Ruocheng Guo, Yushun Dong and Jundong Li 5464

Training-Free Visual Potential Field Calibration for Balanced Hallucination Mitigation in MLLMs
Guangzong Si, Hao Yin, Xianfei Li, Qing Ding, Tao He, Wenlong Liao, Guodong Guo, Pai Peng and Wentao Zhu 5479

REGREACT: Self-Correcting Multi-Agent Pipelines for Structured Regulatory Information Extraction
Mohammed Ali, Abdelrahman Abdallah and Adam Jatowt 5497

PhysDox: Benchmarking LLMs on Physical Feasibility Auditing of Physiological Sensing Protocols
He Liu, Boyuan Gu, Shuaiqi Cheng, Haiyang Sun, Siyu You and Xuming Hu 5539

When Critiques Do Not Change Decisions: DSAR for Threshold-Aligned LLM Reviewers
Zijun Mao, Peng Zhang, Liu Lun and Haozhe Yang 5561

Disagreeing Rationales: Rethinking Classification and Explainability Evaluation in Hate Speech Detection
Benedetta Muscato, Beiduo Chen, Gizem Gezici, Siyao Peng, Barbara Plank and Fosca Giannotti 5586

FaithRL: Learning to Reason Faithfully through Step-Level Faithfulness Maximization
Runquan Gui, Yafu Li, Xiaoye Qu, Ziyan Liu, Yeqiu Chen and Yu Cheng 5603

Weak Ties, Strong Signals: Efficient Training Data Detection in Diffusion LLMs via Independent Token Sampling
Hongyao Yu, Tianqu Zhuang, Ziyuan Xu, Hao Fang, Jiaxin Hong, Bin Chen and Shu-Tao Xia 5631

Beyond the Wrong Bit: Failure States and the Supervision Unit in Verifiable Reasoning
Zihan Ma, Zhuoyi Lin, Minnan Luo, Xiaoxuan Guo and Wenya Wang 5645

From Facts to Insights: A Persona-Driven Dual Memory Framework and Dataset for Role-Playing Agents
Rongsheng Zhang, Ruofan Hu, Weijie Chen, Jiji Tang, Junnan Ren, Wanying Wu, Xunuoyan Chen, Tangjie Lv, Tao Jin and Zhou Zhao 5666

MedCUA-Bench: A Screenshot-Only Benchmark for Clinical Computer-Use Agents
Jia Yu, Zilong Wang, Xinyang Jiang, Dongsheng Li and Shuo Wang 5687

Beyond Text-Intrinsic Features: An Agentic Framework for Evidence-Based AI-Generated Text Detection
Chi Cui, Yixin Wu, Michael Backes and Yang Zhang 5713

Where Does the Signal Live? A Web Data Recipe for Medical Encoder Pretraining
Bofeng Huang, Jacques Sun, Diane Bouchacourt, Nicolas Barascud and Fajwel Fogel 5732

Kinship Data Benchmark for Multi-hop Reasoning
Tianda Sun and Dimitar Kazakov 5767

Cluster-Level Attention-Guided Parallel Decoding for Masked Diffusion Language Models
Heqiang Qi, Wei Huang, Mingyuan Bai and Xiangming Meng 5785

Mind the Perspective: Let's Reason Recursively for Theory of Mind
Chao Lei, Guang Hu, Meng Yang, Yanbei Jiang and Nir Lipovetzky 5806

VERDICT: Verifiable Evolving Reasoning with Directive-Informed Collegial Teams for Legal Judgment Prediction
Hui Liao, Chuan Qin, Yongwen Ren, Hao Li, Zhenya Huang, Yanyong Zhang and Chao Wang 5831

Entropy-Aware Token Rejection for Improving Speculative Decoding
Tiancheng Su, Meicong Zhang and Guoxiu He 5847

Phoneme- and Word-Level Metrics Using Self-Supervised Speech Representations for Forced Alignment Evaluation
V.S.D.S.Mahesh Akavarapu, Michael Daniel and Gerhard Jäger 5865

A Single Suffix to Break Them All: Basin-Aware Jailbreaks for Merged Model Families
YU Zhe, Yixin Tan, Junhao Wei and Chen Wang 5893

Reasoning over Grammar: Can Synthetic Linguistic Reasoning Traces Enhance Low-Resource Machine Translation?
Renhao Pei, Yihong Liu, Sampo Pyysalo, Hinrich Schuetze and Shaoxiong Ji 5916

MAGneT: Coordinated Multi-Agent Generation of Synthetic Multi-Turn Mental Health Counseling Sessions
Aishik Mandal, Tanmoy Chakraborty and Iryna Gurevych 5934

Unveiling Language Routing Isolation in Multilingual MoE Models for Interpretable Subnetwork Adaptation
Kening Zheng, Wei-Chieh Huang, Jiahao Huo, Zhonghao Li, Henry Peng Zou, Yibo Yan, Xin Zou, Jungang Li, Hanrong Zhang, Xuming Hu and Philip S. Yu 5984

CoSpaDi: Compressing LLMs via Calibration-Guided Sparse Dictionary Learning
Denis Makhov, Dmitriy Shopkhoev, Magauiya Zhussip, Ammar Ali and Stamatios Lefkimmiatis 6005

SciOrch: Learning to Orchestrate Expert LLMs for Solving Frontier Multimodal Scientific Reasoning Tasks
Jingru Guo, Xiangyuan Xue, Lian Zhang, Wanghan Xu, Siki Chen, Wanli Ouyang, Lei Bai, Philip Torr and Zhenfei Yin 6032

KFinEval: A Curated Benchmark Suite for Real-World Korean Financial Language Understanding
Seonkyu Lim, Bokwang Hwang, Jonghyuk Baek, Jingu Kim, Hangyeol Yoo and KyungTae Lim 6049

Low-Resource Safety Failures Are Action Failures, Not Representation Failures
Rashad Aziz, Ikhlasul Akmal Hanif and Fajri Koto 6082

EvoHarmBench: Breaking Content Moderation with Iterative Human-Like Evasion
Ruijie Jian, Benlei Cui, Ting Ma, Haidong Ding, Kangwei Liu, Ziwen Xu, Longtao Huang, Hui Xue, Ziqiang Zhu, Junjie Li and Haiwen Hong 6106

PaletteAgent: Self-Evolving Correction for Line-Art Colorization
Binjia Zhou, Guanghui Wang, Hui Shan, Weikai Chen, Zeyu HU, Yingda Yin, Runze Zhang, Kai Yan, Lingting Zhu, Xin Wang and Keyang Luo 6130

Roles with Rails: Contract-Preserving Role Evolution in Multi-Agent Structured Reasoning
Ling-Yue Ge and Lan-Zhe Guo 6159

SparseAlign-OOC: Auditing News Image-Text Out-of-Context Detection with Sparse Autoencoders
Guoxin Wang, Junjie Liu and Shaowu Chen 6193

BASIS: Batchwise Advantage Estimation from Single-Rollout Information Sharing for LLM Reasoning
Shijin Gong, Erhan Xu, Kai Ye, Giulia Livieri, Francesco Quinzan and Chengchun Shi 6208

ReMCTS: Reflect-Enhanced Monte Carlo Tree Search Strategy for Code Generation
Wang Huifei, Huang Xinying, Sun Yiheng and Yifan Yuan 6227

Diagnosing CFG Interpretation in LLMs
Hanqi Li, Lu Chen and Kai Yu 6248

TRACE: A Traceable Agentic Co-Optimizer for Experimental Optimization
Haoran Xu, Keyuan Zhou, Keyan Ding, Jiangzhen Fu, Hui Qian, Qiang Zhang and Huajun Chen 6264

S2M-Inject: Prompt-Guided Zero-Shot Voice Cloning for Music Generation
Teng Ma, Chunyu Qiang, Zhongliang Liu, Xiaopeng Wang, Kang Yin, Ziyu Zhang, Yuzhe Liang, Yushen Chen, Yuxin Guo and Feiyu Chen 6282

Tracking Truth Through Layers: Temporal Dynamics of Fact Representations for Hallucination Detection in LLMs
Yanqiang Zhang, Feng Dawei, Bo Ding, Yuanzhao Zhai and Huaimin Wang 6295

Structural Fabric of Data Shapes the Geometry of LLM Training
Xinyu Lyu, Hanqi Li, Shenghan Zuo, Kai Yu and Lu Chen 6312

HiDiffTIR: Hierarchical Difficulty-Aware Policy Optimization for Multi-Turn Tool-Integrated Reasoning
Yucan Guo, Xiaohan Wang, Miao Su, Saiping Guan, Zhongni Hou, Jiajun Chai, Wei Lin, Guojun Yin, Xiaolong Jin, Jiafeng Guo and Xueqi Cheng 6325

StateSwap: Probing Support–Elimination Hidden States in Multiple-Choice Questions
Chao Gao, Haijiang Liu, Qiyuan Li, Caicai Guo, Frank van Harmelen and Jinguang Gu 6343

Language-Informed Flow Matching for Trend-Guided Structure-Based 3D Molecular Generation
Tianyu Gao, Zhikai Su, Jiashu Li, Wenjun Gao, Ying Zichuan, Zhe Zhao, Fei Zhang and Ye Wei 6365

Representation-based Masked Diffusion Model
Yangrong Hu, Ding Huang, Xueyu Zhou and Jian Huang 6404

Addressing the Reasoning Gap: Mechanistic Circuit-Based Knowledge Editing in Large Language Models
Tianyi Zhao, Yinhan He, Wendy Zheng and Chen Chen 6426

LLMs versus the Halting Problem: Characterizing Program Termination Reasoning
Oren Sultan, Jordi Armengol-Estapé, Pascal Kesseli, Julien Vanegue, Dafna Shahaf, Yossi Adi and Peter O'Hearn 6440

Calibration is the Bottleneck: An Action-Class Diagnostic of Multi-Turn Tool-Calling
Kangjia Zhao, Jiajun Li, Haozhan Shen, Wei Chow, Linfeng Li, Hang Song, Lingdong Kong, Chen Zhi, Tiancheng Zhao, Songhua Liu and Jianwei Yin 6463

Strong Drafts Need Compact Memories: Long-Context Speculative Decoding with Compressed KV Cache
Tong Yuan, Chengxi Liao and Zeyi Wen 6481

NeuronGuard: Robust LLM Safety Alignment via Ablation-Aware Safety Signal Redistribution
Anjun Gao, Yueyang Quan, Yufei Xia, Zhuqing Liu and Minghong Fang 6495

M-VLM2Vec: Structure-Preserving Matryoshka Representation Learning for Vision-Language Embeddings
Minh Khoi Nguyen Do, Nguyen Hong Dang, Danh Vinh Le, Minh Quoc Nhat Nguyen, Linh Ngo Van, Duc Anh Nguyen and Trung Le 6515

Beyond Accuracy: Improving LLMs' Science Communication Skills
Mattan Yeroushalmi, Maxim Bragilovski and Nir Grinberg 6538

GeoGR²: Zero-Shot Geospatial Inference via Geostatistically-Guided Iterative Refinement with LLMs
Jinfan Tang, Kunming Wu, Xieruifeng Gong, Yuya He, Hu Jie, Wu Junhui and Yuankai Wu 6560

Don't Count the Edits, Judge by the Outcome Alone: Reward-Based Evaluation for Grammatical Error Correction
Hayeong Ryu, Sunhee Jo, Seunguk Yu and YoungBin Kim 6581

Balancing Forgetting and Utility with Intervention-Guided Localized Unlearning
Wei Letian, Liangru Xie, Hong-Ning Dai and Hao Wang 6598

Detecting and Explaining Fake News Short Videos with Multimodal Content and Real-World Evidence
Yifeng Luo, Yupeng Li, Ming Tang, Jianxiong Guo and Liang Lan 6614

When and How Should an Agent Clarify? CIGAsk: Teaching LLMs to Clarify via Counterfactual Information Gain
Yunxiang Li, Xixin Wu and Helen M. Meng 6634

APeB: Benchmarking Personalization Ability of Large Language Model Agents
Garry Yang, Zizhe Chen, Xinru Chen, Yongqiang Chen, Jianxiang Wang, Deyu Zou, Linyi Ding, Jialiang Wu, Yunzhong He, Yu Gong, James Cheng and Huaixiao Tou 6647

DoPR: Reusable Compressed Document Prefixes for Efficient LLM Reranking
Beiya Dai, Yifan Wei, Guang Yang, Xing Shi, Xinbing Wang and Zhouhan Lin 6679

FormalEvolve: Neuro-Symbolic Evolutionary Search for Diverse Autoformalization
Haijian Lu, Wei Wang and Jing Liu 6694

SWIM: Student Writing Simulation via Proficiency-Conditioned Generation
Heejin Do, Jakub Kontak and Mrinmaya Sachan 6723

Analyzing Bias in False Refusal Behavior of Large Language Models for Hate Speech Detoxification
Shuzhou Yuan, Kyuri Im and Michael Färber 6741

OmniFusion: Simultaneous Multilingual and Multimodal Translations via Modular Fusion
Sai Koneru, Matthias Huck and Jan Niehues 6770

Exploring Evidence-Grounded Review Generation with Multimodal Inline RAG
Mingrui Zan, Yuxin Guo, Yunquan Zhang, Shijie Ma, Boyang Zhang, Daning Cheng and Fangming Liu 6789

AgendaFlow: Task-Adaptive Meeting Protocols for Multi-Agent LLM Reasoning
Ming Sun, Yin Wang and Yanlin Zhang 6809

Look Before You Leap: Factual Decoding with Internal Attribution Signals
Hayeong Ryu, JungMin Yun, Byeonggeuk Lim, Sunhee Jo and YoungBin Kim 6836

Knowledge Bias in LLM-based Sequential Recommendation: Source, Persistence, and Mitigation
Xiaoyu Dong, Quanyu Dai, Nuo Chen, Qijiong Liu, Yue Zixuan, Xuyang Xie, Jieming Zhu, Zhenhua Dong and Xiao-Ming Wu 6858

Back on Track: Aligning Rewards and States for Reasoning in Diffusion Large Language Models
Yawen Shao, Jie Xiao, Kai Zhu, Yu Liu, Hongchen Luo, Xueyang Fu, Yang Cao, Wei Zhai and Zheng-Jun Zha 6874

Analyzing and Steering Reflection in Large Reasoning Models: A Difficulty-Aware Geometric Approach
Ge Teng, Chen Shen, Ting Mao, Xiang Tian, Xuesong Liu, Rongxin Jiang, Yaowu Chen and Jieping Ye 6891

Towards Omni-dimensional GUI Agent Navigation with Masked Trajectory Prediction
Yan Zhang, Pei Fu, Daiqing Wu, Huawen Shen, Ruoceng Zhang, Shaojie Zhang, Jiahui Yang, Yu Zhou, Can Ma, Zhenbo Luo and Jian Luan 6916

Lost in Choice: Evaluating LLMs Under Extreme Multiple-Choice Scaling
Alessio Cocchieri, Luca Ragazzi, Giuseppe Tagliavini and Gianluca Moro 6931

APEX : Adaptively Balancing Off-Policy Guidance and On-Policy Exploration for LLM Reasoning
Yuchen Wang, Zikang Zhang, Duo Xu, Kunpeng Xu, Qihui Zhu, Wei Huang, Siying Wu, Caihong Cheng, Xin He, Yang Yang, Guanjun Jiang and Xiaoyan Sun 6964

TC-BiMamba: Efficient Dynamic-Chunk Bidirectional Mamba for Unified Streaming and Offline ASR
Jing Peng, Qingshun She, Yangui Fang, Yu Xi, Shuai Wang and Kai Yu 6976

AEScorer: An Agentic Evidence-Grounded Framework for Graded Factuality Verification
Hui Huang, Muyun Yang and Yuki Arase 6986

Communication Alignment in Multi-Agent Systems: From MARL to LLM Agent Communication
Xinghai Wei, Jie Yuan, Hwei-Ming Chung and Tingting Yuan 7005

MEMOBench: A Process Level Memory Benchmark for Robotic Manipulation
Haiyang Sun, Haoxiao Wang, Junming Chen, Weicheng Fang, Zihao Su, Jingkun Yi, Wenyou Yi, Hao Chen and Zhou Zhao 7025

LITERARYBIGFIVE: Author-Personalized Text Generation in a Unified Interpretable Space
Jinghui Zhang, Lang Gao, Ao Li, Mingzhe Li, Ruihong Zeng, Zirui Song, Kentaro Inui and Xiuying Chen 7041

RAGPulse: A Real-World RAG Workload Capturing Authentic Temporal Reuse Patterns for Efficient Serving
Zhengchao Wang, Yitao Hu, Zhuxuan Chang, Jianing Ye, Guotao Yang, Tianhao Lai, Jiazheng Yu, Zhixin Zhao, Liang Zheng, Deng Youpeng, Sheng Chen and Keqiu Li 7065

GroundRAG: Evidence-Verified Multi-Hop Reasoning on Lightweight Graphs
Jinjie Hao, Guixuan Zhang, Ying Huang, Jie Liu and Shuwu Zhang 7077

Detecting and Guiding LLM-Generated Korean Poetry with Interpretable Form-level Features
Keunhyeung Park, Seunguk Yu and YoungBin Kim 7090

LUPE: Long-History User Profile Evolution for LLM-based Recommendation
Yifei Sun, Lu Wang, Fangkai Yang, Pu Zhao, Liqun Li, Mingzheng Li, Jianjin Zhang, Jimeng Sun, Zhiwei Dai, Nan Hu, Zhengxin Zeng, Yuefeng Zhan, Weihao Han, Hao Sun, Qingwei Lin, Saravan Rajmohan, Weiwei Deng, Feng Sun, Qi Zhang and Dongmei Zhang 7110

OptiSet: Unified Optimizing Set Selection and Ranking for Retrieval-Augmented Generation
Yi Jiang, Sendong Zhao, Jianbo Li, Haochun Wang, Bairui Hu, Yanrui Du and Bing Qin . . . 7125

Storyline Trees: Hierarchical Representations for Long-Form Narratives
Litu Ou and Mirella Lapata 7140

RealFin: Benchmarking Agent Workflow Orchestration Beyond MCP-Style Tool Connectivity
Ying Huang, Shuoling Liu, Liyuan Chen, Jie Shi, Jiaqing Liang, Yang Pengtao, Ying Liao, Jinyi Han, Sihang Jiang and Yanghua Xiao 7161

PIVOT: Perplexity-Informed KD-to-RL Transition Scheduling for Vertical-Domain Few-Shot Distillation
Heng Li, Yong Zhang, Ning Cheng, Zhigen Li, Yun Zhu, Yanmeng Wang, Shaojun Wang and Jing Xiao 7184

HALT: Verification-Aware Stopping for Retrieval-Augmented Search Agents
Daeyoung Roh and Donghee Han 7197

Diagnosis Before Treatment? Investigating Intermediate-Variable Reasoning in LLMs
Zhishan Yuan, Tiago Almeida, Shudong Hao, Zining Zhu, Nickolas Elias Dawlabani, Ghazia Pervaiz and Yue Ning 7220

Hierarchical Alignment: Block-Level LoRA-DPO for Preference Tuning in Large Language Models
Yishen Chen, Qi Dong and Yukun Zhang 7236

ESCRAG-R1: Retrieval-Augmented Reinforcement Learning for Emotional Support Conversation
Weichu Liu, Yuxuan Hu, Yirong Sun, Ningning Mao, Ziyun Zhang, Jian Chen, Mingyang Xu, Qishan Zhong and Chengming Li 7253

Associative Memory Updates in Language Models Predict Human Reading Times
Christian Clark and William Schuler 7267

Vis-Poison: Poisoning Visual Knowledge in Multimodal Retrieval-Augmented Generation
Rujin Liang, Zhongpu Chen, Yuhao Lei and Xin Miao 7280

A Survey of Audio Reasoning in Audio Foundation Models
Zhihan Guo, Wenqian Cui, Guan-Ting Lin, Daxin Tan, Jingyao Li, Qiyong Zheng, Dingdong Wang, Jing Xiong, Han Shi, Jiaya Jia and Irwin King 7295

Does Accuracy Equal Evidence? Reasoning Faithfulness under KV Cache Compression
Mengting Ai, Jingrui He and Yue Guo 7313

MemeTrace: A Multimodal LLM-Based Agent for Safety-Aware Meme Provenance Analysis
Bo Shao, Yihan Ma, Yiting Qu, Michael Backes and Yang Zhang 7334

TabEmbed: Benchmarking and Learning Generalist Embeddings for Tabular Understanding
Minjie Qiang, Mingming Zhang, Xiaoyi Bao, Xing Fu, Yu Cheng, Weiqiang Wang, Zhongqing Wang and Ningtao Wang 7355

MAS-Shield: A Defense Framework for Secure and Efficient LLM MAS
Kaixiang Wang, Zhaojiacheng Zhou, Bunyod Suvonov, Jiong Lou, Zihan Wang, Yuxiang Zheng, Yidan Lin, Wutong Zhang, Xianghan Kong, Chentao Wu and Jie LI 7378

BizSage: A Self-Evolving Multi-Agent Framework for Business Research with Efficient Knowledge Retrieval
Yuhe Wu, Guangyu Wang, Jiaxin Liu and Guang Zhang 7399

HEAL: Hindsight Entropy-Assisted Learning for Reasoning Distillation
Wenjing Zhang, Jiangze Yan, Jieyun Huang, Yi Shen, Shuming Shi, Ping Chen, Ning Wang, Zhaoxiang Liu, Kai Wang and Shiguo Lian 7430

Align, Don't Fuse: Structure-Preserving Alignment of Collaborative and Semantic Signals for LLM-Enhanced Recommendation
Jooweon Choi, MiHyeon Kim, Eunju Lee, Junehyoung Kwon and YoungBin Kim 7444

SuperValid: Capability-Aligned OOD Validation for Generalizable Downstream Scaling
Quanen Sun, Changxin Tian, Ke Shi, Cai Chen, Cunyin Peng, Jia Liu, Kunlong Chen and Zhiqiang Zhang 7480

ORBIT: Traceable Rare-Disease Diagnosis via Orphanet-Grounded Retrieval and Bounded LLM Interaction
Songchi Zhou, Haichao Chen and Sheng Yu 7497

ConWriter: Transition-Constrained Stateful Long-Form Story Generation with Lightweight Neuro-Symbolic Consistency Control
Jindong Li, Yang Yang, Zihao Liu, Yutao Yue and Menglin Yang 7515

Dynamically Allocating Evaluation Effort for Model Ranking
Vilém Zouhar, Julia Kreutzer, Alon Lavie, Tom Kocmi, Matt Post, Ondřej Bojar and Mrinmaya Sachan 7533

DyHist: Query-Aware Historical Dynamics Modeling for Temporal Knowledge Graph Extrapolation
Peiyuan Jiang, Yao Liu, Yubo Wen, Yanglei Gan, Lu Liu, Ruojie Zhang, Jiaye Yang, Yuewei Xue, Yajiao Deng, Xiaoyu Guo and Qiao Liu 7557

Same Intent, Different Script: Cross-Script Safety Degradation in LLMs Across Punjabi Sociolinguistic Variants
Arnav Attri and Anuj Attri 7575

CoIn: Counting the Invisible Reasoning Tokens in Commercial Opaque LLM APIs
Guoheng Sun, Ziyao Wang, Bowei Tian, Meng Liu, Zheyu Shen, Shwai He, Yexiao He, Wanghao Ye, Yiting Wang and Ang Li 7594

Revisiting Segmentation in Multi-bit LLM Watermarking
Beining Huang, Youdu Liu, Fei Sun, Du Su, Shuangjie Fu, Huawei Shen and Xueqi Cheng . 7621

See More, Think Deeper: Query-Expanded Visual Evidence and Answer-Clue Guided Reflection for Long Video Understanding
Shuning Wang, Zhiheng Wu, YiNuo LU, Naiming Liu, Chen Jia, Bowen Liu, Shuo Nie, Weijie Zhu and Yumeng Zhang 7643

TxSum: User-Centered Ethereum Transaction Understanding with Micro-Level Semantic Grounding
Zifan Peng, Jingyi Zheng, Yule Liu, Huaiyu Jia, Qiming Ye, Jingyu Liu, Xufeng Yang, Mingchen Li, Qingyuan Gong, Xuechao Wang and Xinlei He 7660

LLM Agents for Time-Series: A Survey
Yilong Chen, Xiao Qin, Chenghao Liu, Liang Wu, Noelle I. Samia and Kaize Ding 7687

DynHD: Hallucination Detection for Diffusion Large Language Models via Denoising Dynamics Deviation Learning
Yanyu Qian, Yue Tan, Yixin Liu, Wang Yu and Shirui Pan 7712

SAVANT: A Neuro-Symbolic Verification Framework for Adjuvant Design
Yi Chen, Yu Zhang, Jian Xu, Hua Yue, Xinming Wang, Zequan Lyu, Boran Wang, Hongyi Liu, Yan Wang, Peiyuan Cao, Xu-Yao Zhang, Wei Wei and Cheng-Lin Liu 7728

ToolSense: A Diagnostic Framework for Auditing Parametric Tool Knowledge in LLMs
Ashutosh Hathidara, Sai Shruthi Sistla, Sebastian Schreiber and Sahil Bansal 7775

GUI-CC: Benchmarking Contextual Consistency of GUI World Models as Agent Environments
Lin Fu, Zheyuan Yang, Tianhui Zhang, Jinbiao Wei, Guo Gan, Boxu Liu, Yilun Zhao and Yu Rong 7798

Caught in the Story: Narrative Captivity in Multi-turn LLMs Conversation
Yuhe Wu, Guangyu Wang, Yujie Chen, Jiatong Zhang, Yuran Chen, Yutong Zhang, Xiyin Cheng, Wenpeng Cao, Zhuang Liu and Guang Zhang 7829

Gradient-Free Test-Time Adaptation for Few-Shot Online Clustering with Large Language Models
Avyav Kumar Singh and Helen Yannakoudakis 7865

Target-Local Verifier Choice in Best-of-K Reasoning Selection
Xin Li, Hao Jiang and Weisi Lin 7892

Syntactic-Guided Semantic Attention Matrix Enhancement for Aspect Sentiment Triplet Extraction
Suhao Yin, Wenyu Guo, Dong Yu, Jian Xue, Zhangfengzhen, Yan Song, Guoqing Jin and Ke Lv 7916

A Man's World: Measuring Gender Bias in Occupational Status Judgments with Vision-Language Models
Ronghao Pan, Tomas Bernal-Beltran, José Antonio García-Díaz and Rafael Valencia-García 7936

CoverPoison: One-to-Many Knowledge Corruption Attacks on Strictly Black-Box RAG Systems
Zihao Yu and Jun Sun 7966

Attributing and Exploiting Safety Vectors through Global Optimization in Large Language Models
Fengheng Chu, Jiahao Chen, Yuhong Wang, Jun Wang, Zhihui Fu, Shouling Ji, Lu Dong and Songze Li 7982

Beyond Preference Learning: Response-Preference Disentangled Reinforcement Optimization for Robust Visual Quality Assessment
Zehui Feng, Weichuan Wang, Xiaohan Chen and Ting Han 7998

Making Every Verified Token Count: Adaptive Verification for MoE Speculative Decoding
Lehan Pan, Ziyang Tao, Xiao Wang, Zhimeng Zhao, Meng Wang and Yanyong Zhang 8025

The Diagnosis Is Already in the Notes: A Channel-Wise Audit of Diagnostic Shortcuts in Clinical-Text Benchmarks
Haocun Ye, Xinlong Jiang, Yuanyuan Hu, Shubai Chen, Tingyu Wu, Zhenkun Zheng, Caijiao Yi, Teng Zhang, Liuxin Wang, Tong Zhao, Heping Xu, Weiwei Dai and Yiqiang Chen 8041

ToSCA: Leveraging Hierarchical Reinforcement Learning on Temporal and Strategic Abstractions of Conversational Agents
Xiaoyu Wang, Qingqing Gu, Yue Zhao, Teng Chen, Yuqi Cao, Xiaokai Chen, Hongyan Li and Luo Ji 8061

Learnable Termination Safeguard: Token-Level Early Termination for Efficient LLM Safety Monitoring
Yulong Xiao, Shanshan Lin, Chao Chen, Yizhe Yang, Xi Zhang and Xiangwen Liao 8084

Entity Blackboard: Structured Agent Memory for Long Document Summarization
Ruyi Liu, Guanyun Zou, Hongliang Dai and Piji Li 8099

EntropyPrune: Matrix Entropy Guided Visual Token Pruning for Multimodal Large Language Models
Yahong Wang, Juncheng Wu, Zhangkai Ni, Chengmei Yang, Yihang Liu, Longzhen Yang, Yuyin Zhou, Ying Wen and Lianghua He 8115

Benchmarking Machine Translation on Chinese Social Media Texts
Kaiyan Zhao, Zheyong Xie, Zhongtao Miao, Xinze Lyu and Shaosheng Cao 8132

EDGE: Error Dependency Graph-Guided Multi-Error Attribution in Multi-Agent LLM Systems
Jun Hou, Priya Pitre, Yi Fang and Xuan Wang 8152

FSA-GRPO: Teaching Auditory LLMs to Use Few-Shot Demonstrations
Haolong Zheng, Siyin Wang, Xulin Fan, Zengrui Jin and Mark A. Hasegawa-Johnson 8192

When Trust Meets Truth: Trust–Truth Separability in LLM-as-Judge
Xin Sun, Di Wu, Yuchen Guo, Jiahuan Pei, Isao Echizen, Abdallah El Ali and Saku Sugawara 8208

Diverse Thinking: Breaking Oppositions in Debates to Foster Synergized Reasoning via Multi-Agent Collaboration
Zhaokang Huang, Yiming Xu, Tao Wang, Yu Yang and Tianyong Hao 8220

Trust Your Guide Only When Certain: Uncertainty-Aware Sparse Alignment at Inference Time
Zeen Zhu, Zhuo Li, Weiyang Guo, Liye Zhao, Haibing Di, Yequan Wang and Jing Li 8237

Aligned but Flattened: Analyzing the Trade-off between Cultural Alignment and Diversity in LLMs
Jingshen Zhang, Shaoyang Xu and Wenxuan Zhang 8264

ClueKV: Self-Guided KV Cache Compression via Adaptive Head Prioritization
Zhuiri Xiao, Jushi Kai and Bo Jiang 8278

OptSkills: Learning Generalizable Optimization Skills from Problem Archetypes via Cluster-Based Distillation
Haochen Yang, Ke Zhao, Mengyuan Ma, Xingyu Lu, Xiangfeng Wang and Hong Qian 8301

MEMMARK: State-Evolution Attribution Watermarking for Agent Long-Term Memory Systems
Haobo Zhang, Xutao Mao, Guangyuan Dong, Ziwei Li, Xuanbo Su, Kaijie Chen, Jing Yang and Zheng Lin 8333

SurveyGraph: Generating Academic Surveys via Methods-Entity Knowledge Graphs
Yifan Yuan, Yancheng Zhu, Zuoyu Zhang and Xiaogang Peng 8350

A Survey of LLM Prompt Datasets: Taxonomy, Linguistic Patterns, and Practical Uses
Yuanming Zhang, Yan Lin, Arijit Khan and Huaiyu Wan 8384

TimelyRAG: Semantic-Temporal Hybrid Retrieval for Time-Critical Question Answering in Overlapping-Evolving Documents
Youngeun Nam, Joeun Kim, Hwanjun Song, Susik Yoon, Jae-Gil Lee and Byung Suk Lee . 8414

MedRAGChecker: Claim-Level Verification for Biomedical Retrieval-Augmented Generation
Yuelyu Ji, Min Gu Kwak, Hang Zhang, Xizhi Wu, Chenyu Li and Yanshan Wang 8431

Learned Sparse Lexical Representations for Large-scale Authorship Attribution
Jian Zhu, Su Yuan and Changbing Yang 8455

Latent Dreaming: Let Language Think Yet Latent See in Driving VLA
Zhixuan You, Ruixu Zhang, Junjie Wang, Yuxin Meng, Lisheng Wu, Lei Zhou, Wei-Bin Kou and Yujiu Yang 8483

Continual Graph Memory for Adaptive Recommendation under Intent Drift
Hao Nguyen Ngoc, Tung Nguyen, Nguyen Thi Hanh, Dinh Thai Hoang and Nguyen Xuan Tung 8501

Rubric-as-Experts: Case-Specific MQM Rubrics for Translation Error Span Detection
Weilu Xu, Yunzhi Shen, Xinye Wang, Renfei Dang and Shujian Huang 8518

SCAMBENCH: A Multi-Dimensional Scam Benchmark through Human-AI Collaborative Data Construction
Huaye Li, Siyu Huang and Jinkyung Katie Park 8540

The Language of Looking: Motion Tokens Reveal Domain-Dependent Local Structure in Gaze Sequences
Virmarie Maquiling and Enkelejda Kasneci 8564

HGP: An on-device personalized agent memory via hybrid graph storage
Ran Zhou, Xueming Han, Jiaheng Liu, Yuyao Zhang, Fanyu Meng, Junlan Feng and Yuxiang Ren 8582

Reverse Probing: Supervised Token-level Uncertainty Quantification for Large Language Models in Clinical Text
Bushi Xiao, Sarvesh Soni and Daisy Zhe Wang 8599

Super Paths, Better Reasoning: Multi-Hop Knowledge Graphs Reasoning with LLM via Super-Path Ranking and Constraint Verification
Kunpeng Du, Yuhang Zheng, Xiaopeng Li, Chen Gao, Jishu Wang, Rui Zhu and Xuan Zhang 8617

Not All Acoustic Frames Contribute Equally: Frame-Selective Acoustic Conditioning for Dysarthric ASR Correction
Gyutae Park and Ji-Hwan Kim 8638

Explication of Refusal: Jailbreaking LLMs by Exploiting Refusal Explanations
Qishen Xu, Ruochun Jin, Ting Wang, Zhongxuan Cai, Ruxin Chen and Huan Chen 8658

When Attention Escapes: Diagnosing and Mitigating Hallucination in Retrieval-Augmented Generation
Wencheng Su, Qinglong Wang, Zhongjie Ba, Zhixuan Chu, Zhenyu Zhong, Liu Yan and Kui Ren 8678

SMRC: Aligning Large Language Models with Student Reasoning for Mathematical Error Correction
BiaoJie Zeng, Min Zhang, Juan Zhou, Fengrui Liu, Ruiyang Huang, YU Song and Xin Lin . 8697

When Knowledge Is Not Free: Cost-Aware Evidence Selection in Retrieval-Augmented Generation
Mingyan Wu, Han Yang, Omer Ben-Porat and Yftah Ziser 8713

Routing Is Not Enough: Diagnosing Intra-Adapter Subspace Contention in MoE+LoRA Fine-Tuning
Mehreen Hossain Chowdhury, Nowshin Mahjabin, Ahmed Shafin Ruhan, MD Azam Hossain, Abu Raihan Mostofa Kamal and Md Tahmid Rahman Laskar 8740

0xClaw: Autonomous Open-ended Agents with Phase-aware Budget Allocation and Temporal Pressure
Yu Zhou, Xinyuan Dou, Yudong Cao and Junhao Song 8753

On the Representation Bias of Disabilities in Generative Text-to-Image Models
Asfandyar Azhar, Zaid Khan, Shaurjya Mandal, Nidhish Shah, Alaa Youssef, Oluwasanmi Adenaiye, Curtis Langlotz and Brad Dicianno 8770

Time Travel Engine: Ordered Temporal Geometry Enables Era-Level Historical Steering in Large Language Models
Jingmin An, Wei Liu, Qian Wang and Fang Fang 8786

CycleIE: Robust Document Information Extraction through Iterative Verification and Refinement
Zhengxuan Zhang, Yin WU, Zhuowen Liang, Haixun Wang, Yuyu Luo and Nan Tang 8813

Coordinate Structure Alignment of Patent Claims by LLMs
Tsukasa Ishimaru, Takehito Utsuro and Masaaki Nagata 8830

Embedding-space Posterior Calibration for Zero-shot Vision-Language Models
Zaiwang Gu, Yingzhi Xia, Yanyu Xu, Xulei Yang and Jun Cheng 8846

Some Emotions Run Deeper: Layer-wise Probing and Causal Intervention in Large Language Models
Tian Fang, Gaël Guibon and Davide Buscaldi 8858

Facts Without Rules: Boundary Metadata Collapse in Multi-Agent LLM Handoffs
Yian Wang, Agam Goyal, Eshwar Chandrasekharan and Hari Sundaram 8877

Video-MME-Logical: A Controlled Diagnostic Benchmark for Video Temporal-Logical Reasoning
Hohin Kwan, Hongyu Li, Renrui Zhang, Manyuan Zhang, Xianghao Kong, Anyi Rao, Jiahao Xie and Si Liu 8900

SagE-QG: Skill-Aligned and Evidence-Grounded Educational Question Generation via Counterfactual Evidence Reinforcement Learning
Yuqiu Deng, Chaoli Zhang, Chenyu Liu, Xingjun Wei, Xinliang Zhou, Zhiguang Wang, Yanfeng Wang, Haoyang Li and Qingsong Wen 8920

RCR-Router:Role-Conditioned Context Routing for Resource-Efficient Multi-Agent LLM Systems
Jun Liu, Pu Zhao, Zhenglun Kong, Wenbin Zhang, Fan Yang, Peiyan Dong, Changdi Yang, Wuzhe Fan, Tianqin Li, Hao Tang, Geng Yuan, Wei Niu, Joannah Nanjekye, Xue Lin, Dong Huang and Yanzhi Wang 8938

Cultural (In)Competence: Investigating Culture-Specific Stereotype Detection Capabilities of LLMs
Maximilian Spliethöver, Jannes Lötfering, Dominik Bachmann, Alexander Espig, Willem Zuidema, Katrin Schulz and Henning Wachsmuth 8964

Two Heads are Better Than One: Test-time Scaling of Multi-agent Collaborative Reasoning
Can Jin, Hongwu Peng, Qixin Zhang, Yang Zhou, Yujin Tang, Tong Che and Dimitris N. Metaxas 8984

Narrowing the Horizon: Quantifying Topic Saliency Shifts in Generative Monoculture
Oriane Peter, Elena Simperl and Kate Devlin 9005

A Skill-augmented Agentic Framework and Benchmark for Multi-Video Understanding
Yue Zhang, Liqiang Jing, Jia Li, Yapeng Tian, Xinya Du, Yunhui Guo and Vibhav Giridhar Gogate 9032

A Model with No Head and Many Thoughts
Nikita Koriagin, Yaroslav Aksenov, George Bredis, Gleb Gerasimov, Nikita Balagansky and Daniil Gavrilov 9058

Regulated Reflection for Efficient Reasoning in Large reasoning language models
Mingqu Wang, Xinjie Xu, Yinan Zou and Qi Ye 9068

ParaShard: Knowledge Sharding for Effective Parametric Retrieval-Augmented Generation
Changjiang Zhou, Ruqing Zhang, Jiafeng Guo, Yingchen Zhang, Fuyu Lv and Xueqi Cheng9084

Beyond Repository Boundaries: Cross-Repository Graph Retrieval for Code Generation
Minh Le-Anh, Nam Le Hai, Quyen Tran, Anh Nguyen Hoang, Linh Ngo Van, Bach Le and Nghi D. Q. Bui 9095

Transformer-Encoder Trees for Efficient Multilingual Machine Translation and Speech Translation
Yiwen Guan and Jacob Whitehill 9115

Beyond Factual Knowledge: Benchmarking and Learning Step-Level Procedural Rule Reasoning in Large Language Models
Bohan Yu, Pengfei Cao, Chen Han, Chenxi Zhou, Zhiheng Zhang, Zhiyang Xie, Wenhao Teng, Xiangwen Liao, Jun Zhao and Kang Liu 9128

DataKernelBench: Can LLMs Optimize Database Queries on GPUs?
Gokul Karthik Kumar, Yotam Perlitz, Corey Lammie, Andrea Giovannini and Katja Hose . . 9153

TRAJDEBUG: Tracing Error Lifecycle to Identify Critical Failures in Long-Horizon Agent Trajectories
Yunjia Qi, Zehua Yin, Xintong Shi, Hao Peng, Songyuanyi Lu, Yixian Liu, Richeng Xuan, Lliu Yuhong, Zhichao Hu, Xiaozhi Wang, Lei Hou, Bin Xu and Juanzi Li 9169

Synthetic Mixed Training: Scaling Parametric Knowledge Acquisition Beyond RAG
Seungju Han, Konwoo Kim, Chanwoo Park, Benjamin Newman, Suhas Kotha, Jaehun Jung, James Zou and Yejin Choi 9193

Self-Evaluation Is Already There: Eliciting Latent Judge Calibration in Base LLMs with Minimal Data
XiuYu Zhang, Yi Shan, Junfeng Fang and Zhenkai Liang 9212

FedLing: Linguistic-Structure-Aware Federated Learning for Fair Multilingual LLM Adaptation
Xiaorui Luo, Hongbo Yin, Ziyu Chen and Yin Zhang 9229

Using Semantic Uncertainty to Estimate Transition Relevance in Turn-Taking
Muhammad Umair and Jan P. de Ruiter 9248

ComplexityMT: Benchmarking the Interaction Between Text Complexity and Machine Translation
Joseph Marvin Imperial, Junhong Liang, Belal Shoer, Abdullah Barayan, Rodrigo Wilkens, Omar Mussa, Dawn Knight, Eugénio Ribeiro, Ekaterina Kochmar, Sowmya Vajjala, Fernando Alva-Manchego and Harish Tayyar Madabushi 9269

Unveiling Sentiment Representations in Code-LLMs
Yongyi Ji, Fuxiang Chen, Jaegul Choo, Seung-Won Hwang and Lu Liu 9284

DVBench: Benchmarking MLLMs for Understanding Dynamic Charts and Narratives in Data Videos
Bomiao Wang, Zekai Shao, Jiexiang Lan, Xiaoliang Fu, Xingchen Zeng and Siming Chen . 9315

DDCG: Decoupled Dual-Critic Guidance for Embodied Agents
Shaojin Ma, Min Zhang, Hongyao Tang, Ruitao Wang, Jianye Hao and Yan Zheng 9339

When Tools Hurt LLM Reasoning: State-Dependent Belief Revision under External Evidence
Yufeng Zhao, Junnan Liu, Hongwei Liu, Dongsheng Zhu, Yuan Shen, Songyang Zhang and Kai Chen 9361

ClawBench: Can AI Agents Complete Everyday Online Tasks?
Yuxuan Zhang, Yubo Wang, Yipeng Zhu, Penghui Du, Junwen Miao, Xuan Lu, Zhuofeng Li, Xingwei Qu, Zhengkang Guo, Yuanzhe Shen, Dingjie Song, Han Zhou, Tuney Zheng, Xian Wu, Hao Yu, Songcheng Cai, Yi Lu, Yunzhuo Hao, Minyi Lei, Liang Chen, Kai Zou, Huifeng Yin, Wendong XU, Dongfu Jiang, Ping Nie, Jiaheng Liu, Wenhu Chen and Kelsey R Allen 9386

RewardCode: Training Generalist Code Reward Model via Pairwise Reinforcement Learning
Songtao Huang, Shiyang Feng, Xiaohan He, Xiangchao Yan, Bo Zhang and Lei Bai 9409

SpeechJBB: Probing Safety Alignment and Comprehension in Large Audio Language Models under Code-Switched Speech
Virginia Ceccatelli, Yejin Jeon and David Ifeoluwa Adelani 9432

E2A-Bench: Benchmarking Evidence-to-Action Reliability in Financial Chart Reasoning
Xiaoya Wang, Yutong Xu and Junjie Wang 9448

HierDiff: Enhancing Large Language Models for Commit Message Generation with Hierarchical Diff Information
Jinrui Sun, Tong Jia, Minghua He and Ying Li 9466

Pruning the Unsurprising: Efficient LLM Reasoning via First-Token Surprisal
Wenhao Zeng, Chao Hu, Yaoning Wang, Yuling Shi, Chengcheng Wan, Hongyu Zhang and Xiaodong Gu 9481

Unifying Conformal Language Tasks with In-Context Ensembles
Xiao Shi Huang, Chen-Yuan Lin, Bruce Kuwahara, Kin Kwan Leung and Jesse C. Cresswell9500

SGA: Plug & Play Geometric Verification for Educational Video Synthesis
Jhon Lopez, Carlos Hinojosa and Bernard Ghanem 9533

ST-Lite: Training-Free KV Cache Compression with Spatio-Trajectory Guidance for Long-Horizon GUI Agents
Bowen Zhou, Zhou Xu, Wanli Li, Jingyu Xiao, Pingan Gan and Haoqian Wang 9550

Not Just Reason, Not Just Scan: Reinforcement Learning for Proactive Scientific Error Verification over Academic Papers
Rongjin Li, Yuanxin Liu, Hao Zhou, Fandong Meng, Jie Zhou and Xu Sun 9587

Demystifying Hidden-State Recurrence: Switchable Latent Reasoning with On-Policy Reinforcement Learning
Jiayu Yang, Chao Chen, Shengen Wu, Yinhong Liu, Yuxuan Fan, Lujundong Li, Songning Lai, Chengwei Qin and Zhijiang Guo 9610

Speculate While You Reason: Teaching Agents to Predict Their Next Tool Call via Joint Agent–Speculator RL
Jiabao Ji, Yujian Liu, Li An, Rohit Jain, Gungor Polatkan, Siyu Zhu and Shiyu Chang 9627

Can Post-Training Turn LLMs into Good Medical Coders? An Empirical Study of Generative ICD Coding
Ziqing Wang, Weihao Li, Shijie Chen, Yuan Luo and Kaize Ding 9643

CHARM: Character Hallucination for Multicultural Role Play Benchmark
Sunkyung Han, Nahyeon Park, Gaeun Seo, Seunghyun Yoon and JinYeong Bak 9659

Compressing Only What Is Compressible: Memory-Efficient Multiscale Muon via Selective Low-Rank Projection
Ren Lianhai and Yeyun Gong 9675

ProActIS: Boosting LLMs' Proactive Information Seeking in Dynamic Clinical Workflows with a Lightweight Exam Proposer
Ruiyang Zhang, Yushi Feng, Liang Peng and Lequan Yu 9693

A Systematic Review of Analogy Generation and Evaluation: Methods, Metrics, and Challenges
Shutong Wu, Tianyu Zhang, Hecong Wang, Yang Lu, Jiankun Yang, Jiamu Tang and Zhen Bai 9710

Mechanistic Interpretability for Understanding Language Abilities in Large Language Models: A Survey
Xufeng Duan, Zhaoqian Yao, Xinyu Zhou, Zihao Fu, Beijia Kan, Yibo Wang, Bei Xiao, Hengyuan Zhang, Xuemei Tang, Ercong Nie and Zhenguang Cai 9735

Which Tokens Should SFT Actually Learn? A Token-Trimming Perspective on Mathematical Reasoning
Yaning Jia, Chunhui Zhang, Wenxuan Xu, Xingjian Diao, Xiaoyuan Wang and Soroush Vosoughi 9760

How Useful are LLMs for Grammar Engineering? Cantonese ParGram Resources and Controlled Experimental Evaluation with English Baselines
Chit-Fung Lam 9774

Text-to-(Single-or-Multi)-SQL: Bridging Real-World Needs with Spider𝕊
Yujian Gan, Xuan Ouyang, Yuxi Lin, Zhilin Zhang, Jinxia Xie and Vassilis Routsis 9792

Don't Just Look, Intervene: Perturbation Based Region Labeling for VQA Images
Marko Jojic, Zhaonan Li and Ben Zhou 9822

Faithful Autoformalization via Roundtrip Verification and Repair
Daneshvar Amrollahi, Jerry Lopez and Clark Barrett 9836

Critic's Eye: Review-Grounded Graph Reasoning for Long-Form Motif Discovery
Xinyu Sun, Zhangyue Yin and Xin Wang 9853

SkillRubricsRL: Evolving Skill-Grounded Agents with Rubric-Guided Reinforcement Learning
Zihe Huang, Jianfei Pan, YiFan Zhang, Tao Chen, Jinghao Lin, Kai Ruan, Xiucheng Xu, Jian Yao, Yijun He and Haihua Yang 9870

Consistency Training Along the Transformer Stack
Sukrati Gautam, Neil Shah, Arav Dhoot, Bryan Maruyama, Caroline Wei, Rohan Kapoor, Robert Sidey, Prakhar Gupta, Zi Cheng Huang and David Demitri Africa 9886

Seeing Before Agreeing: Aligning Multi-Agent Consensus with Visual Evidence
Yuhan Wang, Shuochen Chang, Yalin Feng, Dongsheng Ma, Yuanzi Li, Zhengren Wang, Yinglong Yang, Yufei Chen, Yikang Wang, Shaoxu Sun and Wentao Zhang 9925

Benchmarking Large Language Models for Adaptive Information Acquisition and Diagnosis with Sequentially Revealed Inpatient Records
Huawei Feng, Dongsheng Shi and Linlin Wang 9953

Feedback as a Guide: Enhancing LLM Comprehension of Modern Chinese Poetry
Shanshan Wang, Derek F. Wong, Jiebo Luo, Jingming Yao and Lidia S. Chao 9965

Candidate Pool Design for Safer Preference-Based Post-Training of LLM Companion Chatbots
Yuzhe Bai, Jia Cheng and Yunzhe Liu 9987

On-Policy Distillation Meets Off-Policy GRPO: Training Compact Instruction-Following Rerankers
Vignesh Prabhakar, Jialing Pan and Anil Babu Ankisettipalli 10001

Probing Negation Understanding in Vision-Language Models With Object Attention
Beth Pearson, Michael Wray and Martha Lewis 10028

Do Multimodal LLMs See Before They Read? Diagnosing Contextual Sycophancy
Yi-Cheng Lai and Hen-Hsen Huang 10046

Code-A1: Adversarial Evolving of Code LLM and Test LLM via Reinforcement Learning
Aozhe Wang, Yuchen Yan, Nan Zhou, Zhengxi Lu, Hong Zhao, Yao Fu, Weihao Jiang, Weiming Lu, Jun Xiao, Yueting Zhuang and Yongliang Shen 10068

HI-REASO: A Multi-Level Multi-Faceted Hindi Mathematical Evaluation Suite for Foundation Models
Md Tanzib Hosain, Rajan Das Gupta, Zannatul Fardaush Tripty and Md Kishor Morol 10090

DOG-DPO:Dynamic Optimization in Geometry for Safety Alignment
Yi Nian, Tiankai Yang, Yudi Zhang, QI Pan, Zelong Xu, Shenzhe Zhu, Qingqing Luan, Yue Huang, Xiangliang Zhang and Yue Zhao 10113

SVR-MAD: A Bayesian-Inspired Framework for Posterior-Guided Multi-Agent Debate
Weifan Jiang, Rana Shahout, Minghao Li, Zhenting Qi, Yilun Du, Michael Mitzenmacher and Minlan Yu 10131

HarnessBridge: Learnable Bidirectional Controller for LLM Agent Harness
Xiaoxuan Wang, Haixin Wang, Alexander K Taylor, Jason Cong, Yizhou Sun and Wei Wang 10141

SWE-Edit: Rethinking Code Editing for Efficient SWE-Agent
Yikai Zhang, Jiaxin Pei, Kenan Li, Qirui Jin, Maoquan Wang, Jin Pan, Yu Kang, Shengyu Fu, Elsie Nallipogu, Junjie Hu, Yufan Huang and Zijian Jin 10168

Predicting LLM Training and Inference Runtime: A Survey of Methods, Datasets, and Open Challenges
Kangda Wei, Ziqian Bi, Ashish Molakalapalli, Haonan Duan, Xiaowei Zou, Vincent Du, Alex Carter and Ruihong Huang 10186

FLAME: A New Dataset on FLemish Accounts of Momentary Experiences
Ratna Kandala, Niels Vanhasbroeck and Katie Hoemann 10204

RewardPartial: A Completion Potential Evaluation Framework for Truncated Best-of-N Paradigm
Xin Wang, Yidong Wang, Qiufeng Wang, Bo Jia, Yilin Niu, Jie Tang and Cunxiang Wang 10220

EtCon: Edit then Consolidate for Reliable Knowledge Editing
Ruilin Li, Yibin Wang, Wenhong Zhu, JingHao Zhang, Chenglin Li, Chenliang Li, Junchi Yan and Jiaqi Wang 10235

REDaCT: Collapse-Free Unlearning in Large Reasoning Models
Joshua Maccoby and Wendy Hui Wang 10257

Your Harness is Not Secure: Benchmarking Real-world Threat of Command Line Interface Agent
Weidi Luo, Qiming Zhang, Tianyu Lu, Xiaogeng Liu, Bin Hu, Chiu Hung Chun, Siyuan Ma, Yizhe Zhang, Xusheng Xiao, Yinzhi Cao, Zhen Xiang and Chaowei Xiao 10276

SATS : Scenario-Anchored Topology-Inspired Structural Scoring
Dohyeon Kim and Jihoon Yang 10313

Contextual Causality with Large Language Models: A Survey
Yiheng Zhao, Jun Yan and Chengming Hu 10333

Do LLMs Exhibit Coherent Knowledge Structures in Mathematical Reasoning? A Perspective from Knowledge Space Theory
Peng Cui, Heejin Do and Mrinmaya Sachan 10349

TipCoder: Reinforcement Learning Boosted Test-time Instruction Proposer for Code Generation
Minyu Chen, Sihao Wu, Ling-I Wu, Song Qin, Jingyang Li, Lei Ning, Jianxin Xue and Guoqiang Li 10363

CD^{4}LM: Consistency Distillation and aDaptive Decoding for Diffusion Language Models
Yihao Liang, Ze Wang, Ximeng Sun, Jiang Liu, Jialian Wu, Hao Chen, Xiaodong Yu, Emad Barsoum, Zicheng Liu and Niraj Jha 10378

SSE-Bio: A Structured Self-Evolving Agent with Agentic Retrieval Policy for Multi-Hop Biomedical Reasoning
Zhaohan Meng, Zaiqiao Meng, Siwei Liu, Hao Xu, Ke Yuan and Iadh Ounis 10411

Correctness Forensics for Batch Speculative Decoding: Diagnosing the Ragged Tensor Problem
Ranran Haoran Zhang, Soumik Dey, Ashirbad Mishra, Hansi Wu, Binbin Li and Rui Zhang 10431

SkillForge: Evolving Verifiable Skills for Reinforcement Learning Agents
Shidong Yang, Ziyu Ma, Tongwen Huang, Xucong Wang, Renda Li, Yiming Hu, Yong Wang and Xiangxiang Chu 10448

SHARP: Span-level Hallucination Annotation for Reasoning Paths in Generated Mathematical Solutions
Timofey Garaev, Makar Gorokhovik, Alexey Korotkov, Maksim Savkin, Iurii Vizilter and Andrey Savchenko 10470

Reliable Proof Generation with LLMs via Analogical Retrieval and Symbolic Verification: A Case Study in Euclidean Geometry
Oren Sultan, Eitan Stern and Dafna Shahaf 10488

SEATauBench: Progressively Adapting Tool-Agent-User Evaluation Into Low-Resource Southeast Asian Languages
My Chiffon Nguyen, Aulia Adila, Saksorn Ruangtanusak, Kittiphat Leesombatwathana, Vissuta Gunawan Lim, Patomporn Payoungkhamdee and Samuel Cahyawijaya 10507

When Does Defendant Statement Matter? A Study of Bias and Persuasion in LLM-Simulated Jurors
Cho-Ying Wu 10544

Lingua Franca or Probing Artifact? Rethinking Latent Language in Multilingual LLMs
Deniz Bayazit, Badr AlKhamissi and Antoine Bosselut 10563

Can Large Language Models Uncover the Structure of Social Opinions?
Chongrui Ye, Yuxiang Liu, Haofei Yu, Jiayu Liu and Jiaxuan You 10584

Chain-of-Alpha: Unleashing the Power of Large Language Models for Alpha Mining in Quantitative Trading
Lang Cao 10607

TimlMemBench: A Diagnostic Benchmark for Governed Use of Long-Term Memory in Dialogue Agents
Fangxing Zhou, Yifei Sun, Jiachen Li, Jinyu Xu, Yuhan Wang, Ruixin Li and Qing Xie . . . 10624

Preference Shapes Relevance: Cross-component Hierarchical Semantic Alignment for Personalized Generative Retrieval
Gaoming Zhang, Angqing Jiang, Jianchun Song, Kena Qi, Dayao Chen, Wei Lin and Defu Lian 10639

Not All Subjectivity Is the Same! Defining Desiderata for the Evaluation of Subjectivity in NLP
Urja Khurana, Michiel T. van der Meer, Enrico Liscio, Antske Fokkens and Pradeep K. Murukannaiah 10661

Learning What Matters: End-to-End Adaptive Calibration Data Selection for LLM Pruning
Mengting Ai, Sirui Chen, Xinrui He, Maggie Zarekarizi, Kommy Weldemariam and Jingrui He 10680

Causal Head Gating: Separating Visual Evidence and Knowledge Priors for Hallucination Mitigation in Multimodal LLMs
Jiale Chang and Ying Li 10700

Author-Stigmatising Hazards in LLM-Generated Literary Translation: Evaluation and Mitigation through Danmei-EPV
Yin Zhang and Deblina Bhattacharjee 10715

Exploring Agentic Tool-Calling Decisions via Uncertainty-Aligned Reinforcement Learning
Yijin Zhou, Linqian Zeng, Xiaoya Lu, Wenyuan Xie, Dongrui Liu, Junchi Yan and Jing Shao 10734

LLMail-Inject: A Dataset from a Realistic Adaptive Prompt Injection Challenge
Sahar Abdelnabi, Aideen Fay, Ahmed Salem, Egor Zverev, Kai-Chieh Liao, Chi-Huang Liu, Chun Chih Kuo, Jannis Weigend, Danyael Manlangit, Alex Apostolov, Haris Umair, João Donato, Masayuki Kawakita, Athar Mahboob, Tran Huu Bach, Tsun-Han Chiang, Myeongjin Cho, Hajin Choi, Byeonghyeon Kim, Hyeonjin Lee, Benjamin Pannell, Conor McCauley, Mark Russinovich, Andrew Paverd and Giovanni Cherubin 10757

Do LLMs Change Their Minds Like Humans? Diagnosing Human–LLM Divergence in Single-Turn Persuasion Judgments
Lin Chen, Yitong Chen and Yong Li 10798

Jailbreaking Jailbreaks: A Proactive Defense for LLMs
Weiliang Zhao, Daniel Ben-Levi, Jinjun Peng, Zhou Yu and Junfeng Yang 10816

SlimSearcher: Training Efficiency-Aware Web Agents via Adaptive Reward Gating
Zequn Xie, Junjie Wang, Dan Yang, Jie Feng, Yue Shen, Jian Wang and Jinjie Gu 10838

RuleWeaver: Benchmarking Rule-Centered Scenario Reasoning for Large Language Models
Bohan Yu, Shi-Yang Li, Pengfei Cao, Jun Zhao and Kang Liu 10854

GraphMed-LT: Patient-Specific Graph Memory with Latent Clinical Thought Refinement for Multi-Turn Medical Conversations
Zhaohan Meng, Zaiqiao Meng, Siwei Liu, Hao Xu, Ke Yuan and Iadh Ounis 10879

Optimal Transport Depth Up-Scaling
Mingzi Cao, Xi Wang and Nikolaos Aletras 10898

Do Reasoning Representations Help Humans Evaluate LLM Outputs?
Jaewoo Lim, Sungbok Shin and Sanghyun Hong 10914

DRIP-R: A Benchmark for Decision-Making and Reasoning Under Real-World Policy Ambiguity in the Retail Domain
Hsuvas Borkakoty, Sebastian Pohl, Cheng Wang, Bei Chen and Yufang Hou 10933

MMG2Skill: Can Agents Distill In-the-Wild Guides into Self-Evolving Skills?
Xinyu Che, Junqi Xiong, Yunfei Ge, Xinping Lei, Shihao Li, Hang Yan, Han Li, Yuanxing Zhang, Zhiqi Bai, Jinghua Hao, Ming Sun, Han Li and Jiaheng Liu 10962

RelayLLM: Efficient Reasoning via Collaborative Decoding
Chengsong Huang, Tong Zheng, Langlin Huang, Jinyuan Li, Haolin Liu and Jiaxin Huang 10993

SCARV: Structure-Constrained Aggregation for Stable Sample Ranking in Redundant NLP Datasets
Feiyu Wu, Xu Zheng, Zhuocheng Wang, YIMing Dai and Hui LI 11013

CoVA-SFT: A Large-Scale Dataset for Chain of Visual Abstractions
Tsung-Han Wu, Heekyung Lee, Anya Ji, Haoming Chen, Trevor Darrell, Joseph E. Gonzalez and David M. Chan 11036

When Retain Constraints Conflict: Mitigating Forget–Retain Interference in Tabular Data
Zijie Liu, Jinhao Duan, Bingqi Shang, Xinming An, Sijia Liu and Tianlong Chen 11052

Representation-Aware Unlearning via Activation Signatures: From Suppression to Entity-Signature Attenuation
Syed Naveed Mahmood, Md Rezaur Rahman Bhuiyan, Tasfia Zaman, Jareen Tasneem Khondaker, Md. Sameer Sakib, K. M. Shadman Wadith, Nazia Tasnim and Farig Sadeque 11066

Reasoning Well, Judging Poorly: Theory of Mind Scaffolding Exposes the Limits of LLM Persuasion Judges
Baktash Ansari, Mouly Dewan and Afra Mashhadi 11088

CUADebug: Diagnosing and Repairing Computer-Use Agent Failures
Weijia Zhang, Kunlun Zhu, Zeyi Liu, Yinting Chen, Tianyi Ma, Jiateng Liu, Jiaxun Zhang, Bingxuan Li, Pan Lu, Xiangru Tang, Heng Ji and Jiaxuan You 11110

Thinking effort aligns between humans and reasoning models in abductive reasoning
Henry Arthur 11136

Epistemic Familiarity is Associated With Belief Stability in Large Language Models
Samantha Dies, Courtney Maynard, Germans Savcisens and Tina Eliassi-Rad 11150

Evaluating Criterion-Conditioned Behaviour of Large Language Models in Content Moderation
Danting Zhang, Bei Peng and Robert Loftin 11176

UniDoc-Bench: A Unified Benchmark for Document-Centric Multimodal RAG
Xiangyu Peng, Can Qin, Zeyuan Chen, Ran Xu, Caiming Xiong and Chien-Sheng Wu 11198

Can temporal article-level credibility signals improve domain-level credibility prediction?
Islam Eldifrawi, Shengrui Wang and Amine Trabelsi 11240

GRAIL: Grounded Visual Alignment of Medical Vision-Language Models for Radiology Report Generation
Sarosij Bose, Ravi K. Rajendran, Biplob Debnath, Konstantinos Karydis, Amit Roy-Chowdhury and Srimat Chakradhar 11262

ConstructCIE: A Dataset for Extracting Causal Information from Construction Accident Narratives
Hung Nguyen, Jaehoon Lee, Namgyun Kim and Kuan-Hao Huang 11284

TIA-VL: Diagnosing and Correcting Template-Induced Ambiguity in Medical Vision-Language Decisions
Yifan Luo, Jiaran Shi, Guoming Wang and Siliang Tang 11303

Counterfactual Sensitivity Regularization for Trace-Sensitive Reasoning in Large Language Models
Sanjeda Akter, Ibne Farabi Shihab and Anuj Sharma 11320

Poly-FEVER: A Multilingual Hallucination Detection Benchmark
Hanzhi Zhang, Sumera Anjum, Heng Fan, Weijian Zheng, Yan Huang and Yunhe Feng . . . 11353

ConvDeck: Conversational Paper-to-Slide Generation via Stage-Specific User Feedback
Tarik Can Ozden, Sachidanand VS, Furkan Horoz, Ozgur Kara, Dilek Hakkani-Tür, Junho Kim and James Matthew Rehg 11370

NewsRECON: News article REtrieval for image CONtextualization
Jonathan Tonglet, Iryna Gurevych, Tinne Tuytelaars and Marie-Francine Moens 11410

Finding the Right Evidence: Factor-Guided Coarse-to-Fine Reasoning for Long Videos
Baixuan Xu, Yinyui XU, Tianshi Zheng, Zhaowei Wang, Weiqi Wang, Haochen Shi, Wei Fan, Jiayu Liu, Qing Zong, Xiyu Ren, Xinyu Geng, Zhitao He and Yangqiu Song 11429

Grounded Decoding: Retrieval-Anchored Probability Fusion for Faithful RAG
Ibne Farabi Shihab, Fariya Afrin, Sanjeda Akter and Anuj Sharma 11451

CoLT-Drive: Counterfactual Long-Tail Benchmarking and Knowledge-Preserving Adaptation for Driving Affordance Prediction
Zhengxu Tang, Guofeng Cui, Ziyu Gong, Xiaozhou Zhang, Ruifeng Deng, Chengzhi Qi, Ke Chen, Sachin Patil, Tianjun Xiao, Langechuan Liu and Pichao Wang 11473

AFRILANGTUTOR: Advancing Language Tutoring and Culture Education in Low-Resource Languages with Large Language Models
Tadesse Destaw Belay, Shahriar Kabir Nahin, Israel Abebe Azime, Ocean Monjur, Marek Rei, Chris Biemann, Shamsuddeen Hassan Muhammad, Seid Muhie Yimam and Anshuman Chhabra 11499

Beyond Textual Chain-of-Thought: A Survey on Action-Grounded Reasoning in Autonomous Driving
Zhengxu Tang, Xiaozhou Zhang, Guofeng Cui, Ziyu Gong, Zi Wang, Yunfei Shi, Ruifeng Deng, Chengzhi Qi, Ke Chen, Sachin Patil, Tianjun Xiao, Langechuan Liu and Pichao Wang 11519

PEPPER: Perception-Guided Perturbation for Robust Backdoor Defense in Text-to-Image Diffusion Models
Oscar Chew, Po-Yi Lu, Jayden Lin, Kuan-Hao Huang and Hsuan-Tien Lin 11562

Auditing and Mitigating Paraphrase Fragility in Medical Vision-Language Models
Syed Faizan, Anuradha Chandramohan and Parameswari Krishnamurthy 11575

Addressing the Concept Faithfulness Gap through Probe-Guided Sparse Autoencoders
Lorenzo Malandri, Fabio Mercorio and Antonio Serino 11593

What Models Know, How Well They Know It: Knowledge-Weighted Fine-Tuning for Learning When to Say "I Don't Know"
Joosung Lee, Hwiyeol Jo, Donghyeon Ko, Kyubyung Chae, Cheonbok Park and Jeonghoon Kim 11612

Lingo-Blocksworld: A Dataset for Linguistically-Informed Blocksworld
Melissa Torgbi, Claire Benet Post, Claire Bonial and Harish Tayyar Madabushi 11626

Nonsense Helps: Prompt Space Perturbation Broadens Reasoning Exploration
Langlin Huang, Chengsong Huang, Jinyuan Li, Donghong Cai, Yuyi Yang and Jiaxin Huang 11655

Simulating Democratic Deliberation: Representing Pluralistic Preferences through Electoral Systems in Multi-Agent LLMs
Ryan Faulkner, Stanisław Szufa, Daniel Hoyer, Terry Jingchen Zhang, Roland Bouffanais, Joel Z Leibo and Zhijing Jin 11673

From Coordinates to Candidate Regions: Temporal Change Localization via Region Selection in Remote Sensing Multimodal LLMs
Juwan Chung, Sungjune Park, Yeongyun Kim and Yong Man Ro 11706

Hidden State Poisoning Attacks against Mamba-based Language Models
Alexandre Le Mercier, Chris Develder and Thomas Demeester 11724

Slow to See, Slow to Suppress: Understanding the Effects of Modality in Context-Memory Conflicts
Athulith Paraselli, Tianze Hua and Ellie Pavlick 11751

Graph Serializations for Reasoning LLMs: A Study Across Knowledge Extraction and Structural Reasoning
Julien Yu, Mihaela Bornea, Avi Sil and Katrin Kirchhoff 11779

Revisiting the Effectiveness of LLM Pruning for Test-Time Scaling
Ocean Monjur, Shahriar Kabir Nahin and Anshuman Chhabra 11795

UserHarness: Harnessing User Minds for Stronger Agent Theory-of-Mind
Cheng Qian, Jiayu Liu and Heng Ji 11817

WebXSkill: Skill Learning for Autonomous Web Agents
Zhaoyang Wang, Qianhui Wu, Xuchao Zhang, Chaoyun Zhang, Wenlin Yao, Fazle Elahi Faisal, Baolin Peng, Si Qin, Suman Nath, Qingwei Lin, Chetan Bansal, Dongmei Zhang, Saravan Rajmohan, Jianfeng Gao and Huaxiu Yao 11837

Evaluating the Semantic Specificity of Representation Steering in Language Models
Moy Yuan and Andreas Vlachos 11853

Online Self-Weighted Fine-Tuning
Haiquan Wen, Yiwei He, Bei Peng and Guangliang Cheng 11872

EpiWorld: Grounding LLM Policy Agents in Epidemiological World Models
Zeeshan Memon, Yiqi Su, Kai Shu, Naren Ramakrishnan and Liang Zhao 11885

PhysProver: Advancing Automatic Theorem Proving for Physics
Hanning Zhang, Ruida Wang, Rui Pan, Wenyuan Wang, BingXu Meng and Tong Zhang . . 11907

AIA^2: Attribute-Agnostic Imbalance Augmentation for Subgroup Robustness
Hanshu Rao, Guangzeng Han and Xiaolei Huang 11927

AdaTIR: Adaptive Tool-Integrated Reasoning via Difficulty-Aware Policy Optimization
Zhaiyu Fang and Ruipeng Sun 11945

On the Robustness of Code LLM Guardrails to Code-mixed and Transliterated Inputs
Nishat Raihan, Dipak Meher, Bhuwan Dhingra and Marcos Zampieri 11962

Identifying Student Real Weaknesses: Learnable Error-Aware Distillation for Medical Question Answering
Yonghe Chen, Xiaoyang Meng, Xiaoding Zhou, Wentian Zhang, Ruimeng Zhao, Ying Lian, Guoqiang Chen, Weiyu Zhang and Wenpeng Lu 11982

Personas on Demand: Hypernetwork Synthesis of Unseen Users
Tyson Walsh, Daniel Runfola, Steven Morse and Trenton W. Ford 12005

MemGuard: Preventing Memory Contamination in Long-Term Memory-Augmented Large Language Models
Hyeonjeong Ha, Jeonghwan Kim, Cheng Qian, Jiayu Liu, William M. Campbell, Yue Wu, Yuji Zhang, Kathleen McKeown, Dilek Hakkani-Tür and Heng Ji 12041

EVALUESTEER: Measuring Reward Model Steerability Towards Values and Preferences
Kshitish Ghate, Andy Liu, Devansh Jain, Taylor Sorensen, Atoosa Kasirzadeh, Aylin Caliskan, Mona T. Diab and Maarten Sap 12072

Dataset Protection via Watermarked Canaries in Retrieval-Augmented LLMs
Yepeng Liu, Xuandong Zhao, Dawn Song and Yuheng Bu 12100

Think-Probe-Respond: Improving Large Language Models as Judges of Research Idea Novelty
Tim Schopf, Tobias Schreieder and Akiko Aizawa 12119

CARE: Privacy-Compliant Agentic Reasoning with Evidence Discordance
Haochen Liu, Weien Li, Rui Song, Zeyu Li, Chun Jason Xue, Xiao-Yang Liu, Sam Nallaperuma-Herzberg, Xue Liu and Ye Yuan 12133

BELIEFMEM: Belief-Revising Memory for Conversational LLM Agents
Felix Meng, Junda Wu and Julian McAuley 12152

Large Language Models for Machine Translation Quality Annotation: Humans and Models Are Both Challenged
Hala Almaghout, Christian Federmann and Qin Gao 12167

Latent Recurrent Thoughts: Recurrent Refinement of Proposed Latents for Reasoning with Frozen LLMs
Zhaoliang Chen and Jie Fu 12185

FormalMATH: Benchmarking Formal Mathematical Reasoning of Large Language Models
Zhouliang Yu, Ruotian Peng, Keyi Ding, Yizhe Li, Zhongyuan Peng, Minghao Liu, Yifan Zhang, Zheng Yuan, Huajian Xin, Wenhao Huang, Yandong Wen, Ge Zhang and Weiyang Liu 12204

Can Aha Moments be Fake? Towards Quantifying Decorative and True Thinking in Chain-of-Thought
Jiachen Zhao, Yiyou Sun, Weiyan Shi and Dawn Song 12242

SynAV: Integrated Grounding and Captioning for Fine-grained Omni Video Understanding via Balanced Audio-Visual Synergy
Bolun Zhang, Guangming Yao, Canjun Yang and Yi Yuan 12259

WebWeaver: Style-Aware Webpage Generation with Style Contracts
Runchen Xu, Mengxiao Zhang, Yikai Gu and Jiamou Liu 12283

The Environmental Impacts of Language Model Training Keep Rising Now is the Time to Catch Impacts on the Rebound
Clément Morand, Aurélie Névéol and Anne-Laure Ligozat 12320

From Monolithic Prioritization to Stratified Replay: Hierarchical Prioritized Experience Replay for Task-Oriented Dialogue Policy Learning
Yixin Zhang, Yangyang Zhao, Kai Xu, Lingyun Xiang and Yun Song 12336

Leveraging Teacher-Forcing Token-Level Anomaly Signals for Heterogeneous Backdoor Filtering in LLMs
Tianxu Han, Jiangming Wang, Chunyi Zhou, Binbin Zhao and Tianyu Du 12351

Seeing the Poem: Image-Semantic Detection of AI-Generated Modern Chinese Poetry with MLLMs
Shanshan Wang, Fengying Ye, Hanjia Lyu, Caiwen Gou, Junchao Wu, Jingming Yao, Chengzhong Xu, Jiebo Luo and Derek F. Wong 12371

Semantic Shifts of Nationalist Lexemes in German: From Pre-War to Post-War Periods (1918–2024)
Yiting Tong, Anni Wang, Yaxi Zhuang and Wei Zhao 12388

UniD^3: Open Knowledge Graphs and QA Benchmarks for Drug–Disease Discovery and Reasoning
Qing Wang, Tianshi Liu, Minghao Zhou, Jialu Liang, Sen Guo, Jing Su and Qianqian Song 12399

Information-Guided Frontier Decoding: Contextual Utility-Driven Commitment in dMLLMs
Xingyou Fang, Jingxing Zhong, Xiaosong Yuan and Xiaofeng Zhang 12422

To Nuke or Not to Nuke: LLMs' (Missing) Ethical Reasoning and Authorization Actions in a High-Stakes Decision-Making Simulation
John Chen, Sihan Cheng, Can Gurkan and H M Abdul Fattah 12436

RL from Teacher-Model Refinement: Gradual Imitation Learning for Machine Translation
Dongyub Jude Lee, Zhenyi Ye and Pengcheng He 12473

DrugClaw and DrugAudit: A Primary-Source-Grounded Agent and Authority-Aware Benchmark for Drug-Information Question Answering
Qing Wang, Bo Li, Jialu Liang, Daling Shi, Bob Zhang and Qianqian Song 12486

BIRD-History: A Benchmark for History-Driven Text-to-SQL with Fine-Grained Knowledge Annotations
Yunfan Zhou, Qiming Shi, Yizhou Yang, Di Weng and Yingcai Wu 12511

When Retrieval Similarity Hurts: A Diagnostic Framework for Document Weighting in Black-Box RAG
Dongju Seo and Beakcheol Jang 12534

Emotional Labor Strategy Preferences in LLM Personas
Mohammad Saim and Tianyu Jiang 12559

Beyond Accuracy: Evaluating Grounded Visual Evidence in Thinking with Images
Xuzhao Li, Xuchen Li, Renjie Pi, Shiyu Hu, Jian Zhao and Jiahui Gao 12577

Can LLMs Reliably Identify Causal Conclusions in Scientific Literature? A Multidisciplinary Evaluation
Young Soo Ko, Nadia Sharief and Xuan Lu 12592

AgentRx: Diagnosing AI Agent Failures from Execution Trajectories
Shraddha Barke, Arnav Goyal, Alind Khare, Avaljot Singh, Suman Nath and Chetan Bansal 12610

MCTS-Debater: Monte Carlo Tree Search based Rebuttal Generation for Competitive Debate
Xiaoyong Huang, Yuting Huang, Guanlan Yue, Hao Wang, Heli Sun, Jiaruo Wu, Zejun Wang and Liang He 12640

FHE-CodeEval: Benchmarking Large Language Models for Fully Homomorphic Encryption Code Generation
Qiuchu Yu, Guangli Li, Shuoming Zhang, Ruiyuan Xu, Qihan Cai, Jingling Xue, Xiaobing Feng and Huimin Cui 12657

Backfire-R1: Identifying and Mitigating Persuasion Vulnerabilities in LLM Agents
Jingwei Sun, Yi Yu, Jingjing Qu, Hanxi Zhu, Jia Xu and Jing Shao 12674

REMP: A Structured Event-Probing Benchmark for Long-Horizon Role-Playing Dialogue
Qianxi Yu, Yuhao Dan, Qingchun Bai, Qin Chen, Jie Zhou and Liang He 12702

Fine-grained stance cues and commonsense knowledge Integrated Reasoning Enhancement for Zero-shot Conversational Stance Detection
Weiming Yin, Jinzhong Ning, Yijia Zhang and Yuanyuan Sun 12723

When Safety Routing Breaks: Understanding Alignment Fragility under Benign Fine-Tuning
Yitong Guo, Xiaoyi Chen, Siyuan Zhang, XiaoFeng Wang and Haixu Tang 12748

Mind the Gap: Theory-of-Mind-Grounded Friction for Epistemic Alignment
Yifan Zhu, Kyeongmin Rim and James Pustejovsky 12763

PRISM: Post-Training LLM Pruning via Residual Importance Signal Merging and Intra-Block Sparsity Allocation
Keyu Wen, Jianing Zhu, Kaiyang HU, Lijun Yang and Rui Lin 12779

VocMoE: Vocabulary-Guided Mixture-of-Experts for Korean Medical Language Models
Sangyeon Cho, Jaeho Han, Mingyu Jeon and Junyeong Kim 12796

Breaking the Structural Identity: Personalized Federated LoRA Fine-tuning under Rank Heterogeneity
Lei Wang, Jieming Bian, Letian Zhang and Jie Xu 12816

BitNet Text Embeddings
Zhen Li, Xin Huang, Liang Wang, Nan Yang, Ting Song, Yan Xia, Xun Wu, Shaohan Huang, Huishuai Zhang, Furu Wei and Dongyan Zhao 12833

MIRAGE: Stealthy Visual Prompt Injection for Vulnerability Detection in Web Agents
Xuelong Dai, Jianyu Ma, Boyang Ma, Biwei Yan, Yijun Yang and Yue Zhang 12850

Non-binary bottom-up constituency parsing without arity actions
Jungyeul Park, Eunkyul Leah Jo and Zihao Huang 12869

NaviRank: LLM-Informed Latent Space Navigation Enhances Reasoning-Intensive Information Retrieval
Jinwen Pan, Haolin Jiang and Hang Liu 12880

Beyond Single Vector: Query-Aware Multi-Perspective Document Embeddings for Retrieval
Zhen Li, Chongyang Tao, Yihan Chen, Huishuai Zhang and Dongyan Zhao 12898

Stop Rolling the Dice: Active Group Relative Policy Optimization via Intrinsic Metacognition
Zheng Gong, Ying Sun, Hao Peng and Philip S. Yu 12915

PCoAD: The Pharmaceutical Certificate of Analysis Dataset Based on LLMs Validation and Calibration
Yiming Xu, Zhaokang Huang, Haoran Yu, Fu Lee Wang and Tianyong Hao 12928

Seeing Failure In Advance: Pre-Generation Failure Diagnosis for Multi-Modal Retrieval-Augmented-Generation
Chan-Wei Hu, Shuo Xing, Ryan A. Rossi and Zhengzhong Tu 12959

DPRES: Dual-Perspective Decoupled Representation Learning Framework for Multi-Trait Essay Scoring
Jiong Wang, Hao Wu and Shengquan Yu 12978

From Benchmarking to Failure Diagnosis: Diagnosing Classical Chinese Understanding in LLMs via IRT and Error Attribution
Jiong Wang, Tengju Li and Shengquan Yu 12991

GMTRouter: Personalized LLM Router over Multi-turn User Interactions
Yihang Sun, Encheng Xie, Tao Feng and Jiaxuan You 13006

ProRet-VLA: Bridging Prospection and Retrospection for Long-Horizon Robotic Manipulation
Enhua Cheng, Xiaoyuan Fang, Yuxuan Wang and Piji Li 13029

Diagnosing and Mitigating Evidence Misgrounding in Multimodal Knowledge Editing
Zhiwu Wang, Yichen Wu, Jinqian Yang, Yizhou Yang, Renzhen Wang, Qian Zhao and Deyu Meng 13048

BackTrend: Evaluating Scientific Weak-Signal Prediction via Backward Reconstruction
Xiao Zhou, Yilun Zhao, Owen Jiang, Tiansheng Hu, Cai Xu, Manasi Patwardhan and Arman Cohan 13065

SAIGuard: Communication-State Simulation for Proactive Defense of LLM Multi-Agent Systems
Ruxue Shi, Yili Wang, Mengnan Du, Qinggang Zhang, Rui Miao, Yixin Liu and Xin Wang 13095

Latent Personal Memory: Represent personal memory as dynamic soft prompts
Debrup Das, Avinash Amballa, Yashas Malur Saidutta, Vijay Srinivasan, Vivek Kulkarni and Srinivas Chappidi 13113

Role Consistency in Role-Playing Agents: A Survey of Evaluation, Memory, Reasoning, Alignment, and Safety
Chenghao Li, Xueqin Xiao, Yinuo Xiao, Zihan Xu, Xinzhao Li, Yuxuan Luan, Jiwei Wei, Yang Yang and Chaoning Zhang 13129

Exploring LLMs for South Asian Music Understanding and Generation
Faria Binte Kader, Mohtasim Hadi Rafi, Shah Wasif Sazzad and Santu Karmaker 13146

Targeted Decoding Intervention for Persona Inconsistency Hallucination Mitigation in Multi-Turn Dialogue
Jie Bao, Yan Xu, Haojie Wu, Yu Gao, Xinyu Cui, Jian Zhao, Yong Liao and Zhen Chen . . . 13164

RxnExBench: Benchmarking Vision-Language Models for Structured Chemical Reaction Extraction from Scientific Literature
Jiaxing Sun, Junyuan Gao, Ruikang Zhang, Hao Zheng, Lingli Ge, Xiao An, Jiang Wu, Huazheng Zeng, Jiahe Song, Chen Zhu, Haote Yang, Jingchao Wang, Lu Chen, Yuanyuan Cao, Ruijie Zhang, Shengxin Lu, Weijia Li, Lijun Wu and Conghui He 13180

FinGuard: Detecting Financial Regulatory Non-Compliance in LLM Interactions
Huaixia Dou, Jie Zhu, Minghao Wu, Shuo Jiang, Junhui Li, Lifan Guo, Feng Chen and Chi Zhang 13213

Temporal Tree of Thought: Reasoning-Guided Visual Cue Search for Long-Video Understanding
Ziling Huang and Shin'ichi Satoh 13233

Sequential Beats Joint: On the Interplay between On-Policy Distillation and RLVR
Boyan Li, Bingsen Chen, Chenghao Yang, Ping Nie, Chen Zhao and Xi Ye 13252

ETR: Outcome-Guided Elastic Trust Regions for Policy Optimization
Shijie Zhang, Kevin Zhang, Zheyuan Gu, Xiang Guo, Rujun Guo, Liushaoyu, Guanjun Jiang and Wangxiaozhao 13271

WhiFlash: Accelerating Speculative Decoding with Token-Level Cross-Paradigm Routing
Young D. Kwon, Miles Williams, Rui Li, Alexandros Kouris and Stylianos I. Venieris 13286

BlueprintAgent: Constraint-Triggered Targeted Revisits for Simulation-Ready Generation from Scanned Structural Blueprints
Zhouyuan Xu, Chen Yang, Linhao Wang, Jiansheng Fan and Chen Wang 13304

Beyond the Sampled Token: Preserving Candidate Support in RLVR
Ruotian Peng, Yi Ren, Zhouliang Yu, Weiyang Liu and Yandong Wen 13318

ARC: An Audit-Guided Risk Calibration Attack for Jailbreaking Multimodal Large Language Models
Wenluo Hu, Chunyi Zhou, Tianyu Du and Shouling Ji 13341

STAMP: Provenance-Guided Credit Assignment for Deep Search Agents
Ke Xu, Han Xu, Xinran Chen, Yuqian Wang, Zhixuan Li, Xiaojian Liu, Changwo Wu, Jianqiang Xia and Yuchen Li 13360

Transfer Safety Awareness for Cross-Modal Safety Drift in Multimodal Large Language Models
Tianqi Xiao, Shiyao Cui, Minghao Zhang, Junxiao Yang and Renmiao Chen 13375

Beyond GRPO and On-Policy Distillation: An Empirical Sparse-to-Dense Reward Principle for Language-Model Post-Training
Yuanda Xu, Hejian Sang, Zhengze Zhou, Ran He, Zhipeng Wang and Alborz Geramifard . 13392

SmartThinker: Learning High-Density Reasoning via Step-Level Token Allocation
Xingyang He, Jie Liu and Xiao Ling 13409

AgenticRag-R1: Agentic Reinforcement Learning with Stack Memory for Multi-Step Reasoning, Retrieval and Memorizing
Xinke Jiang, Yue Fang, Zhibang Yang, Jiaran Gao, Zhixin Zhang, Tao Feng, Rihong Qiu, Wentao Zhang, Hongxin Ding, Ruizhe Zhang, Yongxin Xu, Yuheng Huang, Xu Chu, Junfeng Zhao and Yasha Wang 13429

TESS: LLM-guided Translation, Exploration, Simulation, and Search for Embodied Task Planning
Peiran Xu, Yixiang Yang, Jinghan Li, Fei Li, Cao Sheng, JiaZhong Yu and Yadong MU . . . 13457

When LLM Meets Tree Search: A Systematic View of Inference as Search in Large Language Models
Jiaqi Wei, Xiang Zhang, Yuejin Yang, Wenxuan Huang, Juntai Cao, Sheng Xu, Xiang Zhuang, Zhangyang Gao, Muhammad Abdul-Mageed, Laks V. S. Lakshmanan, Chenyu You, Wanli Ouyang and Siqi Sun 13488

OmniHallu: Unified Hallucination Detection for Cross-Modal Comprehension and Generation in Multimodal Large Language Models
Jianjiang Yang, LI Peihang, Shanqing Xu, Mengchen Qian, Lu Zhang and Meng Luo 13544

FLOW-ABSTAIN: Modeling the Answerable Manifold for Query-Aware Abstention in Large Language Models
Jixin Zhang and Bensheng Qiu 13556

Decomposition-Guided Diffusion Language Models for Inertial Confinement Fusion Prediction
Xiang Zhang, Varchas Gopalaswamy, Rahman Ejaz, Riccardo Betti and Dongfang Liu . . . 13567

COMPASS: Cognitive MCTS-Guided Process Alignment for Safe Search Agents
Wenkai Shen, Pengyang Zhou, Jiahe Xu, Jiaming Qian, Haozhe He, Zhihao Huang, Chaochao Chen and Xiaolin Zheng 13583

RepairGate: RL-Trained Skill Repair and Abstention for Lifelong Agent Libraries
Xiaoyu Chen, Fengge Wu, Junsuo Zhao and Yun Fan 13610

BrandEval: A Two-Track Multi-Agent Benchmark for Risk-Sensitive LLM Crisis Communication
Haotian Gan, Weidong Tang, Hanbin Sun, Yi Luo, Xintian Li, Jiayong Wang, Wanyue Li and Yudong Li 13631

little m: An AI Agent for Industrial Process Optimization
Yongchao Ye, Xinyu he, Dutliff Boshoff, Way Kuo and Lishuai Li 13658

TempCloze: Can Video-LLMs Identify the Missing Middle?
Wenqi Pei, Henry Hengyuan Zhao, Yilai Liu, Jiahao Meng, Chen Han, Ziyu Wang and Hongyang Du 13676

CEDAR: Automata as Verifiable Interfaces for Language-Guided Embodied Action
Lekai Chen, Alvaro Velasquez and Ashutosh Trivedi 13694

RIFT: Relation Modeling and Transfer for Visual Figurative Entailment
Xiaojian Lou, Qimeng Yang, Hui Song, Yunjing Jiang, Jinmiao Song and Xinjun Pei 13711

How Much Do LLMs Know About Chinese Zero Pronouns?
Yifei Li, Guanyi Chen and Tingting He 13726

Latent Reward Steering: An Adaptive Inference-Time Framework that Implicitly Promotes Cognitive Behaviors in Reasoning LLMs
Jiakang Li, Guanyu Zhu, Can Jin, Chenxi Huang, Dexu Yu, Ronghao Chen, Yang Zhou, Hongwu Peng, Xuanqi Lan, Dimitris N. Metaxas and Youhua Li 13748

HintSteer: Predicting Context Imprints for Inference-Time Steering
Xinda Wang, ChenZhuo Zhao, You Hu, Guotong Li, Chuang Ou and Yuhao Zheng 13768

SinkPruner: Sink-Free Visual Token Pruning for Multimodal Large Language Models
Shiyu Li, Zi-Yuan Hu, Shijia Huang, Yanyang Li, Yiwu Zhong and Liwei Wang 13776

Reframe to Retrieve: Cognition-aware RAG with Projections Inspired by Appraisal and CBT for Conversational Mental Health Support
Minh-Tien Nguyen, Huu-Loi Le, Manh-Cuong Phan, Duong-Hieu Le, Dinh-Phuc Pham, Van-Quyet Nguyen and Hajime Hotta 13800

Fluid Reasoning Representations
Dmitrii Kharlapenko, Terry Jingchen Zhang, Arth Singh, Alessandro Stolfo, Arthur Conmy, Mrinmaya Sachan and Zhijing Jin 13823

CLIMB: Confidence-Guided Complementary Evidence for Multimodal Retrieval-Augmented Generation
Hang Gao, Wujiang Xu, Zhixing Zhang, Kai Mei, Jingyi Yang and Dimitris N. Metaxas . . 13842

SPARK: Skeleton-Guided Reasoning Synthesis from Large-Scale Scientific Literature
Yu Li, Wei Li, Xin Gao, Mengyuan Sun, Xiaoyang Wang, Qizhi Pei and Lijun Wu 13858

SwarmBench: Can Large Language Models Act as Agent Swarm Orchestrators?
Jinshan Gao, Zhuoran Jin, Tianyi Men, Kang Liu and Jun Zhao 13882

Beyond Individual Personas: Aligning Synthetic Dialogue to Population-Level Behavior Distributions
Xinyi Liu, Rinat Khaziev, Hooshang Nayyeri, Emine Yilmaz, Charith Peris and Hari P Thadakamalla 13903

TARQ: Tail-Aware Reconstruction Quantization for Rare-Word Robust Automatic Speech Recognition
Xinyu Wang, Ziyu Zhao, Ke Bai, Silin Meng, Dongming Shen, Xiao-Wen Chang and Yixuan He 13928

StatefulDiscovery: Evidence-Calibrated Claim Formation in Open-Ended Scientific Discovery
Jiayao Chen, Shi Liu and Linyi Yang 13953

Toward Robust Personalized Alignment for LLMs: Mitigating Persona Drift in Multi-Turn Dialogue
Youyuan Zhang, Siyuan Li, Fangming Liu and Jing Li 13978

UniBrowse: A Data-to-Agent Framework for Multimodal BrowseComp
Xiyu Wei, Qingwei Zong, Zhuocheng Yu, Feng Zhang, Kai Jia and Sujian Li 14007

Aligning Agentic World Models via Knowledgeable Experience Learning
Baochang Ren, Yunzhi Yao, Rui Sun, Shuofei Qiao, Ningyu Zhang and Huajun Chen 14024

Compiler-Guided Adaptive Proof Search with Cross-Model Synergy on Context-Dependent Theorem Proving
Zhuo Liu, Ding Yu and Hangfeng He 14055

Target-Aware Calibration Data Selection for Preserving Uncertainty in Quantized Language Models
Zhen Yang, Sizai Hou, Kaiwen Zheng, Yaofang Liu, Liang He, Yixuan Chen and Kangning Cui 14073

Joint Training of Multi-Token Prediction in Reinforcement Learning via Optimal Coefficient Calibration
Zili Wang, Jiajun Chai, Lin Chen, Xiaohan Wang, Shiming Xiang and Guojun Yin 14094

STOP: Structured On-Policy Pruning of Long-Form Reasoning in Low-Data Regimes
Chenjun Xu, Zhennan Zhou, Zhan Su, Bill Howe, Lucy Lu Wang and Bingbing Wen 14109

ThinkFlow: Self-Evolving Probabilistic Latent Memory for Lifelong Conversational Agents
Cai Ke, Xin Liu, Han Zhang, Jiangyue Yan, Zike Yuan, Ling Deng, Yue Yu, Hui Wang and Ruifeng Xu 14126

OViP: Online Vision-Language Preference Learning with Hallucination-Targeted Negatives
Shujun Liu, Siyuan Wang, Zejun Li, Jianxiang Wang, Zeng Cheng, Xuanjing Huang and Zhongyu Wei 14142

FinBalance: A Multi-Document Accounting Reconciliation Benchmark
Sasank Tumpati, Devansh Agarwal, Ayush Kedia, Arjun Neekhra, Murari Mandal, Krishna Garg, Yash Sinha, Suman Gupta and Dhruv Kumar 14168

Tool-meme: Tool-Learning Framework for Multimodal Meme Detection
Shuai Liu and Yingying Yao 14192

Continuous Difficulty Control for Text Simplification
Yuki Fujiwara, Rina Miyata, Tomoyuki Kajiwara and Yuki Arase 14216

Ban&Pick: Enhancing Performance and Efficiency of MoE-LLMs via Smarter Routing
Yuanteng Chen, Peisong Wang, Yuantian Shao, Nanxin Zeng, Chang Xu and Jian Cheng . . 14228

Sparse Model Diffing via Dynamic Circuits
Yueyan Li, Wenhao Gao, Caixia Yuan and Xiaojie Wang 14254

Atomicity for Agents: Exposing, Exploiting, and Mitigating TOCTOU Vulnerabilities in Browser-Use Agents
Linxi Jiang, Zhijie Liu, Haotian Luo and Zhiqiang Lin 14277

Stage-Aware Controllable Dialogue Generation for CBT-Based Psychological Support
Haibo Chen, Qiuyang Zhao, Qinyuan Liu, Bo Jiang, Meng Sun, Mingmei Gao and Qingwei Chong 14292

RGCC: Retriever-Guided Context Compression for Efficient Vision–Language Alignment in Knowledge-Based VQA
Qiankun Li, Feifei Zhai, Jianghao Chen, Chong Li, Haoran Gao, Fanyu Meng, Junlan Feng, Yang Zhao and Yu Zhou 14313

Interactive Memory Learning for Long-Term Conversations
Cai Ke, Jiangyue Yan, Han Zhang, Xin Liu, Zike Yuan, Yue Yu, Hui Wang and Ruifeng Xu14336

Adaptive Bi-SADA: Bidirectional Structure-Aware Data Augmentation for Robust Aspect Sentiment Quadruplet Prediction
Feixiang Liu, Qiang Qiu, Hui Xu, Zheng Wang and Xueqi Cheng 14359

Atoms and Grammar of the Heart: Geometry-Aligned Discrete Tokenization for ECG Language Modeling
Zirui Wang, Qiuling Yang, Pang Wu, Peng Wang, Zhenfeng Li, Xianxiang Chen, Lidong Du and Zhen Fang 14369

Evidence-Bounded Mental Health Reasoning from Heterogeneous Speech Protocols
Chengyuan Gao, Jiang WU, Tao Lu, Jiayan Guo, Mingkun Xu, Tianyi Zang and Shangyang Li 14394

CVaR-RMD: CVaR Selection Pressure for Robust Metaheuristic Discovery
Ruodai Cui and Li Niu 14407

Escaping the Frequent Finding Shortcut: Pathology-Aware Curriculum Learning for Radiology Report Generation
Ziming Song, Mingyang Shen and Chi Man Vong 14430

DynSess: Dynamic Session-Level Evaluation and Optimization Framework for Role-Playing Agents
Rongsheng Zhang, Jiji Tang, Junnan Ren, Zuyi Bao, Weijie Chen, Ruofan Hu, Tangjie Lv, Zhou Zhao and Yan Zhang 14443

LaRA: Layer-wise Representation Analysis for Detecting Data Contamination in RL Post-Training
Minju Gwak, Minseo Kwak, Dongseok Lee, Guijin Son, Alan Ritter and Jaehyung Kim . . 14462

Monitorable Chart Reasoning Agents via Verifiable Process Rewards
Sanchit Sinha, Oana Frunza, Kashif Rasul and Aidong Zhang 14488

OptiArena: Can LLMs Improve Executable Algorithms under Fixed Resource Budgets?
Wenjun Peng and Xinyu Wang 14512

Same Semantics, Different Outcome: On the Modality Robustness of Multimodal LLMs under Knowledge Conflict
Jungyeon Lee, Yejin Yoon and Taeuk Kim 14529

Beyond Semantic Similarity: Learning Behavior-Aligned Retrieval for Tool-Augmented LLMs
Yixin Chen, Ying Xiong, Shangyu Wu, Yufei Cui, Xue Liu, Nan Guan and Chun Jason Xue 14549

ARIA: Audio-Driven Melody–Tone Relation Modeling for Cantonese Lyric Authoring
Shengyu Li, Jinting Wang and Li Liu 14572

UR^2-MLLM: Uncertainty-aware Revisit Reasoning in Multimodal Large Language Models for Radiology Report Generation
Yucheng Chen, Yang Yu, Jiazhou Zhou, Yufei Shi, Yongying Lan, Yichi Zhang, Li Liyi and Si Yong Yeo 14596

Small Language Models as Judges for Rubric-Based Reinforcement Learning
Fengyu Xie, Yilun Zhao, Bingsen Chen, Arman Cohan and Chen Zhao 14621

InterventionBench: Can LLM Agents Adapt to Mid-task User Interventions?
Jaehoon Lee, Suyoung Bae, Jin-Seop Lee, YunSeok Choi and Jee-Hyong Lee 14646

STAR : Sentence Translation Alignment Rate for Document-to-Document Machine Translation
Yichen Dong, Hao Wang, Junhui Li, Linlong Xu, Longyue Wang and Weihua Luo 14667

PREFINFER: Preference Inference from User-LLM Interaction History
Yufei Chen, Hao Peng, Yunjia Qi, Lei Hou, Bin Xu and Juanzi Li 14690

ForeSight: Enhancing Risk Monitoring via Early Safety Signal Distillation
Hanling Wang, Chenlong Wei, Ling Xu, Hanyan Niu, Qi Cao, Shizhou Huang, Yang Yang, Xiaohui Zhu and Yao Zhu 14713

FIVER: Adaptive Fusion of Formal and Informal Verification for Mathematical Reasoning in LLMs
Yujie Li, Ao Xu, Ziyou Guo and Tieru Wu 14730

Entropy-Adaptive Fine-Tuning: Resolving Confident Conflicts to Mitigate Forgetting
Muxi Diao, Lele Yang, Wuxuan Gong, Yutong Zhang, Zhonghao Yan, Yufei Han, Kongming Liang, Weiran Xu and Zhanyu Ma 14751

EviRerank: Adaptive Evidence Construction for Long-Document LLM Reranking
Minghan Li, Eric Gaussier, Juntao Li and Guodong Zhou 14772

NOVA: NOise-aware Verbal Confidence CAlibration for Robust Large Language Models in RAG Systems
Jiayu Liu, Rui Wang, Qing Zong, Yumeng Wang, Cheng Qian, Qingcheng Zeng, Tianshi Zheng, Haochen Shi, Dadi Guo, Baixuan Xu, Chunyang Li and Yangqiu Song 14793

ConTExT: Temporal Knowledge Graph Question Answering via Contractive Dual-Space Subgraph Search
Xu Quan, Yanting Yin, Shiying Li and Wensheng Zhang 14836

DeployBench : Benchmarking LLM Agents for Research Artifact Deployment
Yuanli Wang, Yaoyao Qian, Yue Zhang, Hanhan Zhou, Jindan Huang, Tianfu Fu, Qiuyang Mang, Huanzhi Mao, Wenhao Chai, Wendong Fan and Liqiang Jing 14848

BelaMorphLM: A Controlled Study of Morphology-Aware Tokenization for Belarusian Encoder Language Models
Dmitry Morozov, Lizaveta Astapenka and Olga Shcherbakova 14867

Don't Wait to Reply: Towards Responsive yet Thoughtful Dialogue through Proactive Thinking
Ante Wang, Jiaqi Fu, Xuanyi Chen, Ruotian Ma, Zhaopeng Tu, Weizhi Ma and Yang Liu . 14886

Perish or Flourish? A Holistic Evaluation of Large Language Models for Code Generation in Functional Programming
Lang Hong Nguyet Anh, Eric Lang, Bach Le, Huynh Quyet Thang and Thanh Le-Cong . . 14903

Feedback-Assisted Trust Propagation over Document Relation Graphs for Retrieval-Augmented Generation
Zhuoheng Li and Ying Chen 14926

ClueAegis: Heuristic-to-Reasoning Cognitive-skill Learning for Unified Evidence-based Synthetic Image Detection
Huangsen Cao, Hongkang Chu, Yuxi Li, Ying Zhang, Chen Li, Jing Lyu, Yongwei Wang, Yu Zhao and Fei Wu 14941

DIP: Dynamic In-Context Planner For Diffusion Language Models
Yang Li, Han Meng, Chenan Wang, Zhenyu Bi, Xuan Wang and Haipeng Chen 14973

FairSynth: Instrument-Relative Evaluation and Calibration of LLM Survey Simulations for Measurement Fairness
Guangrui Fan, DanDan Liu, Qian Guo, Xueying Niu and Lihu Pan 14992

Searching for a Unified Approach to Morpheme Segmentation: Benchmarking Supervised and LLM Architectures Across 20 Languages
Dmitry Morozov, Anna Glazkova and Olga Lyashevskaya 15013

Stabilizing Hard-Negative Preference Optimization for Medical LLMs
Kazuma Kobayashi, Yosuke Yamagishi, Ryota Shibaki, Azusa Sakai, Takashi Kodama, Lin Gu, Irene Li, Yuki Arase, Akiko Aizawa and Sadao Kurohashi 15038

EvoSQL: Memory-Augmented Critic-Generator Co-Evolution for Text-to-SQL
Jiawei Zhou, Jianwei Wang, Chenyu Zhou, Chaojian Shi, Ming Dong and Kai Wang 15064

Grounded Normative Rule Generation with Structured Search
Fanqi Kong, Huaxiao Yin, Ruijie Zhang, Xiaoyuan Zhang, Yizhe Huang, Jian Gao, Shuo Chen and Song-Chun Zhu 15093

What to Keep in Tight-Budget Dialogue Compression: Turn-Level Information Matters
Minji Park, Seunghyun Yoon and Hyuk Lim 15123

GeoMesh: Workload-Balanced and Sign-Compressed Geo-Distributed LLM Training
Changyong Shin, Jaerim Park, Minchul Kang, Younghun Go, Zhixiong Niu, Yongqiang Xiong, Gyeongsik Yang and Chuck Yoo 15131

Unifying Detection and Adaptation in Task-Free Continual Learning
Dezheng Han, Anbang Zhang, Zhihao Zhu and Shuaishuai Guo 15153

From Parameters to Behaviors: A Survey of Model Fusion for Large Language Models
Shuo Cai, Yanggan Gu, Zihao Wang, Yuanyi Wang, Yibo Yan, Wenjun Wang, Yuhang Liu, Guanghao Zhu, Sirui Huang, Li Ming and Hongxia Yang 15178

VDAR-Router: Adaptive LLMs Routing via Verbalized Query Difficulty Analysis Retrieval
Yu-Chien Tang, Jun-Chen Hung, Wen-Chih Peng and An-Zi Yen 15203

FailAlign: Failure-Aligned Bootstrapping for Long-Horizon HTML Based Code Generation Agents
Fanglin Xu, Wei Zhang, Jian Yang, Guo Chen, Liqun Yang and Zhoujun Li 15223

Can Language Model Agents be Helpful Circuit Explainers in Mechanistic Interpretability?
Ayan Antik Khan, Harsh Kohli, Yuekun Yao, Huan Sun and Ziyu Yao 15237

TrustFV: Trustworthy Fact Verification on Knowledge Graph via Process Reward-Guided Step-Level Constrained Decoding
Yue Zhang, Shicheng Zhou, Yifu Gao, Zhiliang Tian, Wenqing Hou and Bin Zhou 15262

Unlocking Latent Medical Reasoning in LLMs via Inference-Time Representation and Prefix Interventions
Ming Ma, Jiayun Li, Sendong Zhao, Yanrui Du, Yuzheng Fan, Danyang Zhao, Bing Qin and Ting Liu 15276

The Illusion of Visual Tool-Use: A Causal Audit of Thinking with Images
Zhiheng Wang, Bo Peng, Lai Wei and Chaochao Lu 15298

Task-Grounded and EL2N-Informed RAG Switch for Domain-Specific Information Extraction
Xingyu Zhu, Robert Bossy, Balazs Nagy, Laszlo Vidacs and Claire Nédellec 15319

HumanAgencyBench: Scalable Evaluation of Human Agency Support in AI Assistants
Benjamin Sturgeon, Daniel Samuelson, Jacob Haimes and Jacy Reese Anthis 15337

CoLa-ICD: A Knowledge-Enhanced Framework for Long-Tail Automated Medical Coding
Yihang Cheng, Veronica Liesaputra and Andrew Trotman 15368

Demo-ICL: In-Context Learning for Procedural Video Knowledge Acquisition
Yuhao Dong, Shulin Tian, Shuai Liu, Shuangrui Ding, Yuhang Zang, Xiaoyi Dong, Yuhang Cao, Jiaqi Wang and Ziwei Liu 15381

MONA: Muon Optimizer with Nesterov Acceleration for Scalable Language Model Training
Jiacheng Li, Jianchao Tan, Hongtao Xu, Jiaqi Zhang, Yifan Lu, Yerui Sun, Yuchen Xie and Xunliang Cai 15400

Information-Theoretic Limits of Positional Identifiability under Hybrid Graph Encodings
Zimo Yan, Zheng Xie, Chang Liu, Yiqin Lv, Runfan Duan, Yuan Wang and Bingyao Zeng 15414

Agentic Chain-of-Thought Steering for Efficient and Controllable LLM Reasoning
Yu Xia, Zhouhang Xie, Xin Xu, Byungkyu Kang, Prarit Lamba, Xiang Gao and Julian McAuley 15435

An Empirical Study on Zero-Data Bootstrapping for Conversational Recommender Systems
Rohan Surana, Junda Wu, Zhouhang Xie, Yu Xia, Nathan Kallus and Julian McAuley 15455

Event Ontology Expansion via LLM-Based Conceptualization
Weicheng Ren, Zixuan Li, Long Bai, Xiaolong Jin, Jiafeng Guo and Xueqi Cheng 15473

Linear Fitness Subspace in Protein Language Models Enables Sample-Efficient Directed Evolution
SiYuan Ma, Canran Xiao, Zikai Xiao, Albert Gao, Liang He, Xuan-Yu Wang, Shuying Cao and Xiaojun Jia 15490

Analyzing Reasoning Shifts in Audio Deepfake Detection Under Adversarial Attacks
Binh Nguyen, Charles Fleming and Thai Le 15512

RealWorldShop: Benchmarking and Improving Conversational Shopping Agents in Real-World E-commerce
Xinwei Yang, Kelong Mao, Yudong Guo, Sulong Xu, Gu Simiu, Chen Huang and Wenqiang Lei 15533

PRAXIS: A 68-Industry Benchmark for Pragmatic-Form Bias in Suggestion Mining
Aniket Upadhyay, Aakash Trivedi, Sparsh Bajoria, Shreyanshu Ranjan, Dhruv Kumar and Pratik Narang 15566

Spectral Query-Key Product Weight Steering for Training-Free VLM Hallucination Mitigation
Karn Tiwari, Varnith Chordia and Prathosh AP 15587

EvoPool: Evolutionary Programmatic Annotation for Label-Efficient Specialized Supervision
Tianyi Xu, Yaolun Zhang, Xuan Ouyang and Huazheng Wang 15617

Doc2Weight: In-Weights Retrieval-Augmented Generation via Parameter Injection
GyuSik Jang, Wonkee Lee and Seung-Hoon Na 15657

From Memory to Behavior: A Behavior-Aware Role-Playing Framework for Social Media Influencers
Ji-Lun Peng, Yi-Zhen Zhang, Chun-Nan Chou and Yun-Nung Chen 15675

When Evidence Conflicts: Reliability-aware Meta-review Generation
Xinzhe Wang, Fei Tao, Jiang Xie, Hong Yu and Ye Wang 15708

More Evidence Is Not Always More Support: Controlled Diagnostics for Evidence Aggregation in Fact Verification
Shaoxin Zhong, Yuchen Su, Qian Liu and Michael J. Witbrock 15724

PRISM: Promoting Sampling Diversity of Large Language Models with Implicit Stochastic Modulation
Qixin Xu, Zihan Zhu, Xiaojiang Liu and Bohan Lyu 15748

DarterSD: Dynamic Token-Weighted Distillation for High-Efficient Speculative Decoding
Zhaoning Zhang, Xu Baizhou, Libo Zhang, Yi Chen, Zhiliang Tian, Zhen Huang and Dongsheng Li 15773

AgentSwing: Adaptive Parallel Context Management Routing for Long-Horizon Web Agents
Zhaopeng Feng, Liangcai Su, Zhen Zhang, Xinyu Wang, Xiaobin Wang, Runnan Fang, Xiaotian Zhang, Hui Chen, Qi Zhang, Joey Tianyi Zhou, Pengjun Xie, Chenxiong Qian, Bryan Hooi and Zuozhu Liu 15789

NS-Copilot: An LLM-Driven Agent System for Autonomous Neuroscience Analysis
Wuche Liu, Yiran Qiao, Linlin Hou, Rui Yang, Shusen Pu, Song Wang and Jing Ma 15809

You Won't Believe This Click: Content Rewriting for Agentic Choice
Tianyi Jin, Zirui Wang and David M. Chan 15829

MANGO: Manga Active Narrative Grounding Optimization
Hao Qiu, Junyan Wang, Zheyuan Liu, Lei Fan, Hong Jia, Lianbo Guo and Zhulin Tao 15858

TSA-RAG: Structure-Preserving Retrieval-Augmented Generation via Tiered Document Graph
Kaiyuan Zhang, Chong Feng, Ge Shi, Lingnan Xu and Bo Wang 15874

OneVision: An End-to-End Generative Framework for Multi-view E-commerce Vision Search
Zexin Zheng, Huangyu Dai, Lingtao Mao, Xinyu Sun, Zihan Liang, Ben Chen and Chenyi Lei 15895

ShotFinder: Imagination-Driven Open-Domain Video Shot Retrieval via Web Search
Tao Yu, Haopeng Jin, Hao Wang, Shenghua Chai, Yujia Yang, Zhongtian Luo, Junhao Gong, Jiaming Guo, Minghui Zhang, Xinlong Chen, Zhenghao Zhang, Yuxuan Zhou, Yanpei Gong, Yuan-Cheng Liu, Yiming Ding, Kangwei Zeng, Pengfei Yang, Yufei Xiong, Shanbin Zhang, Chengshao-xiong, Huang Ruilin, Li Shuo, Yuxi Niu, Xinyuan Zhang, Yueya Xu, Jie Mao, Ruixuan Ji, Yaru Zhao, Mingchen Zhang, Jiabing Yang, Jiaqi Liu, YiFan Zhang, Hongzhu Yi, Xinming Wang, Cheng Zhong, Xiao Ma, Zhang Zhang, Yan Huang and Liang Wang 15915

P-Space: Puncturing the Null-Space to Resolve Cognitive Dissonance in Continual Knowledge Editing
Shiqiang Tian, Cheng Ding, Qin Chen, Jie Zhou, Xin Li and Liang He 15945

CEE: An Inference-Time Jailbreak Defense for Embodied Intelligence via Subspace Concept Rotation
Jirui Yang, Zheyu Lin, Zhihui Lu, Yinggui Wang, Lei Wang, Tao Wei, Qiang Duan, Xin Du and Shuhan Yang 15965

SkillComposer: Learning to Evolve Agent Skills for Specification and Generalization
Qi Zhang, Zhaopeng Feng, Xiaonan Shi, Xiaomeng Hu, Liuchu, Pengjun Xie, Xiaobin Wang, Bryan Hooi, Haobo Wang and Junbo Zhao 15997

Search-Time Contamination in Deep Research Agents: Measuring Performance Inflation in Public Benchmark Evaluation
Yongjie Wang, Xinyue Zhang, Yao Kunhong, Zhiwei Zeng, Kaisong Song, Jun Lin and Zhiqi Shen 16012

When Can Large Reasoning Models Save Thinking? Mechanistic Analysis of Behavioral Divergence in Reasoning
Rongzhi Zhu, Yi Liu, Jiancheng Wang, Xiangyu Liu, Zequn Sun, Yiwei Wang, Yu Deng, Zijian Zhou and Wei Hu 16028

MedMT-Bench: Can LLMs Memorize and Understand Long Multi-Turn Conversations in Medical Scenarios?
Lin Yang, Yuancheng Yang, Xu Wang, Changkun Liu and Haihua Yang 16041

When Seeing Is Not Believing—A Benchmark for Search-Grounded Video Misinformation Detection
Tao Yu, Yujia Yang, Shenghua Chai, Jinshuai Zhang, Haopeng Jin, Hao Wang, Minghui Zhang, Zhongtian Luo, Yuchen Long, Xinlong Chen, Jiabing Yang, Zhaolu Kang, Yuxuan Zhou, Siyi Cao, Yuqing Wu, Peifan Wang, Tianyang Dang, Minghui Guo, Hanyu Wang, Hai Huang, Zhengyu Man, Xingyu Yang, Xinya Sun, Xinming Wang, Hongzhu Yi, Zheqi He, Xi Yang, Yan Huang and Liang Wang 16076

Attributing Structured-Output Gains in Function Calling: Interface Alignment versus Procedural Transfer
Wanyi Chen, Daoyuan Chen and Fang Kong 16119

Attention-Aware Remasking: A Training-Free Approach for Enhancing Visual Fidelity in Multimodal Diffusion Language Models
Yechan Cha and Dong-Kyu Chae 16137

Agentic Reward Modeling for Medical Reasoning
Hang Zhang, Ruheng Wang, Yuelyu Ji, Min Gu Kwak, Xizhi Wu, Chenyu Li, Li Zhang, Wenqi Shi, Yifan Peng and Yanshan Wang 16152

When Less Is More: An Empirical Study of Prompt-Pattern Composition
Xiang Li, Tongqing Shi, Zhirui Li, Jiale Han and Benyou Wang 16169

GRAD: GRADient Rectification based on Authentic Data for Low-Resource Neural Machine Translation with Back Translation
Zhuofei Liang, Xinglin Lyu, Lingling Mu and Hongfei Xu 16189

OASIS: An Open-vocabulary Benchmark for Aspect-Opinion-Sentiment Extraction with Implicit-aspect Specificity
Junyeop Im, Jonghyun Park and Dong-Kyu Chae 16201

A Benchmark for Self-Evolving Agents via Experience-Driven Lifelong Learning
Yuxuan Cai, Yipeng Hao, Jie Zhou, Hang Yan, Zhikai Lei, Rui Zheng, Zhenhua Han, Yutao Yang, Junsong Li, Qianjun Pan, Tianyu Huai, Qin Chen, Xin Li, Kai Chen, Bo Zhang, Xipeng Qiu and Liang He 16232

InsightChain: Optimized Chain-of-Insight Analytics for LLM-driven Data Visualization
Sun Hanya, Chen Zhang, Sheng Liang, Yongyue Zhang and Yong Liu 16273

ClinicalReTrial: Clinical Trial Redesign with Self-Evolving Agents
Sixue Xing, Kerui Wu, Xuanye Xia, Haoyu He, Meng Jiang, Jintai Chen and Tianfan Fu . . 16296

GSSPO: Group Sentence-Specified Sequence Policy Optimization for LLM Personality Alignment
Shurui Zhang and Zikai Li 16331

Towards Effective Experiential Learning: Dual Guidance for Utilization and Internalization
Fei Bai, Zhipeng Chen, Chuan Hao, Ming Yang, Ran Tao, Bryan Dai, Xin Zhao, Jian Yang and Hongteng Xu 16363

Ask Only When Needed: Proactive Retrieval from Memory and Skills for Experience-Driven Lifelong Agents
Yuxuan Cai, Wei Li, Jie Zhou, Qin Chen, Xin Li, Bo Zhang and Liang He 16384

Citation-Closure Retrieval and Per-Rule Attribution for Real-World Regulatory Compliance Question Answering
Yeong-Joon Ju and Seong-Whan Lee 16401

OVIBench: Benchmarking Online Video Question Answering under Interruption
Naiming Liu, Zhiheng Wu, Shuning Wang, Tie Zhang, Bowen Liu and Tong Wang 16427

Disentangling In-Language Alignment and Cross-Lingual Safety Borrowing in Low-Resource Languages
Zeguang Huang, Zhenma Spenpa, Mio Yip Sien Chu, Xi Cao, Nuo Qun and Nyima Tashi . 16447

MemFail: Stress-Testing Failure Modes of LLM Memory Systems
Ishir Garg, Neel Kolhe, Dawn Song and Xuandong Zhao 16460

PATS: Process-Level Adaptive Thinking Mode Switching
Yi Wang, Junxiao Liu, Shimao Zhang, Jiajun Chen and Shujian Huang 16503

The Age of Curiosity Meets the Age of AI: Benchmarking Child Safety in Large Language Models
Samee Arif, Angana Borah and Rada Mihalcea 16520

Beyond Reflection: Affirmation as a Promising Behavioral Marker Associated with Quality in Text-Based Counseling
Michimasa Inaba 16557

Auditing and Mitigating Privacy Leakage in Cloud-Edge Collaborative Decoding
Kejia Zhang, Tianyuan Zou, Zixuan GU and Yang Liu 16572

RoutePO: Personalized Routing Policy Optimization for LLM-based Multi-Channel Recall
Zaiyi Zheng, Yaochen Zhu, Mingxuan Ju, Tong Zhao, Neil Shah and Jundong Li 16593

PrefTool: Lightweight Preference-Aware Retrieval for Personalized Tool Selection
Xuanzhen Li and Linan Yue 16610

Mitigating Exploration Bias in RL for Multi-Instruction Following
Mian Zhang, Yueqin Yin, Kaiyu He, Peilin Wu, Xinlu Zhang, Mingyuan Zhou and Zhiyu Chen 16624

SCAFFOLD: Self-Improving Web Agents via Recursive Parametric Skill Abstraction
Bowei He, Xiaokun Zhang, Meng Ding and Xue Liu 16634

How Order-Sensitive Are LLMs? OrderProbe for Deterministic Structural Reconstruction
Zhaolu Kang, Yingjie He, Kehan Jiang, Leqi Zheng, Jiachen Qian, Qianyuan Zhang, Chunlei Meng, Yujie Feng, Yuan Wang, Stephen Dou, Aming Wu, Pengxiang Zhao, Jiaxin Liu, Guansu Wang, Zeyu Zhang, Zeyu Wang, Lei Wang, Qishi Zhan, Xiaomin He, Meisheng Zhang, Jianyuan Ni and Richeng Xuan 16651

EIBench: A Simulator-Based Benchmark and Turn-Credit RL for Emotion Management
Rongzhi Zhu, Xiang Huang, Yuchuan Wu, Rui Wang, Zequn Sun, Tao Ren, Weiyao Luo, Bingxue Qiu, Jieping Ye, Yongbin Li and Wei Hu 16682

CoRe-MoE: Compact Reusable MoE for Continual Multimodal Instruction Tuning
Runze Liu, Naibin Gu, Mingxu Ai, Yuqing Li, Peng Fu, Zheng Lin and Weiping Wang . . . 16699

Modality Fault Lines: Structural Corruptions Reveal Fragile Omni-Modal Reasoning
Zhaolu Kang, Meixin Wu, Yu Xue, Yingjie He, Qiming Shi, Lei Wei, Yidi Wang, Richeng Xuan and Zhichao Hu 16715

Diagnosing Order Dependence in Vision-Language Task-Progress Judgment with Equal-State Supervision
Can Wang, Xiaoda Yang, Xueyi Zhang, Menglan Tang, Jingyang Xue, Shenzhou Gao, Jiahe Zhang, Xiaoran Geng, Jiangning Zhang, Hongyu Yuan, Peijian Zhang, Xiaobin Hu and Jianwen Lou 16749

RegioZH: Unveiling Comprehension Disparities across Regional Chinese Varieties in LLMs
Weihua Zheng, Wu Kui, Thong T. Doan, Lin Jingxia, AiTi Aw, Nancy F. Chen and Roy Ka-Wei Lee 16774

Cascade: Hierarchical Recoverability Control for Large Language Model Unlearning
Qingchen Yu, Shiying Duan, Xiaodong Li, Yuhua Wang, Zhiyu li, Shiji Zhou, Yifan Sun and Zhaoxin Fan 16812

PeLA: Prompt-Enhanced Logit Adaptation for Personality-Trait Steering in LLMs
Haoxuan Kou, Wenxuan Wang, Zirui Liu and Jiaxing Shen 16839

Evidence Absence Is Not Evidence Insufficiency: Diagnosing NEI Construction Artifacts in Fact Verification
Jingxi Qiu, Zeyu Han and Cheng Huang 16862

CoEffi: Enhance Efficient Code Generation via Dynamic Collaborative Decoding
Jiahui Lin, Jingrui Zhang, Zixuan Zhang, Meilin Li, Man Fu, Siyao Yu, Weifeng Sun and Meng Yan 16886

A Cross-lingual Comparison of Human and Classification Model Entrainment Behavior in Code-switched Speech Settings
Debasmita Bhattacharya, Siying Ding, Alayna Nguyen and Julia Hirschberg 16906

KV Cache Quantization for Diffusion Language Models
Yan Tai, Yunan Ding, Guangzhao Li, Siqi Luo, Chaoyang Zhang, Guangtao Zhai, Guodong Guo and Xiaohong Liu 16937

Improving Information Extraction with Learned Queries
Omar Sharif, Soroush Vosoughi and Nikhil Singh 16952

Beyond Single Gold Labels: Human Variability and Conceptualisation in Chinese Classifier Use
Yongheng Peng, Guanyi Chen and Tingting He 16973

From Topical Relevance to Answerability: Entailment Distillation for Conversational Retrieval
Shuai Qin, Guojia An, Weikang Guo, Pei Ke, Jiwei Wei, Yang Yang and Jie Zou 16989

DualEval: Joint Model-Item Calibration for Unified LLM Evaluation
Aaron J. Li, Hao Huang, Youngmin Park, Yitong Ma, Wei-Lin Chiang, Li Chen, Cho-Jui Hsieh, Bin Yu and Ion Stoica 17004

PaperGraph-Bench: Graph-Guided Benchmark Construction and Multi-Turn Evaluation for Scientific Paper Reading
Rao Fu, Qing Dai, Kaitai Ou-yang, Kang He, Fei Li, Donghong Ji and Chong Teng 17028

Know When to Reflect: Entropy-Guided Adaptive Self-Correction for LLM Reasoning
Xinglei Hao, Haiyu Zhao, Chunhong Zhang, Liping Liu, Jiarun Liu and Zheng Hu 17056

Take Only What You Need: Rank Minimization as an Implicit Forgetting Regularizer in Continual Learning
Haodong Lu, Chongyang Zhao, Jason Xue, Lina Yao, Kristen Moore and Dong Gong 17078

The Illusion of What If: Evaluating the Breakdown of Counterfactual Reasoning in LLMs
Yucheng Wang, Yuetian Du, Zhengyi Liu, Rongyu Zhang, Bing Zhao, Boyu Yang, Ming Kong, Lin Qu, hu Wei, Jie Liu and Qiang Zhu 17106

IterInject: Indirect Prompt Injection Against LLM Agents via Feedback-Guided Iterative Optimization
Zixuan Chen, Jiaxiang Chen, Li Luo, Ke Xu, Xiaoxiang Huang, Tanfeng Sun and Xinghao Jiang 17130

Typology-Aware Dual-Track: Detecting Hybrid AI Text Across Languages
Kihoon Park, Seungwoo Lim and Jihie Kim 17150

LLM Pedagogical Behavior in AI Tutoring Interactions
Suhyeon Lee, Juneha Baek, Jaehyeong Park and Donghyuk Shin 17168

Style-Ruler: Inference-Only Stylized Translation via Unsupervised Extraction and Agentic Collaboration
Hong Ren, Jiahe Li, Shaolin Zhu, Deyi Xiong and Nannan Wu 17178

Diagnosing Subword Boundary Conflicts in Diffusion Language Models through Sentence Simplification
Yaowen Xu and Ming Cai 17196

RouteVerify: A Diagnostic Study of Tool-Augmented Chest X-ray Agents
Haochen Liu, Yukuo Li, Xia Liu, Qingwei Wang, Shourong Xin, Zhuohui Liu and Long Zhao 17213

ARIA: Affect-Responsive Interaction Alignment for Emotional Support Policies via Implicit Multimodal Affective Feedback
Deuksin Kwon, Kevin Joo, Elena Hayoung Lee, Mohammad Soleymani and Gale M. Lucas17226

Learning Solvability Boundaries in LLMs: Improving Reasoning, Detecting Unsolvability, and Calibrating Refusal
Dengyun Peng, Qiguang Chen, Bofei Liu, Jiannan Guan, Libo Qin, Zheng Yan, Jinhao Liu, Jianshu Zhang and Wanxiang Che 17250

Decoupled Analysis-Judging: An Automated Creativity Evaluator Using LLMs in Complex Multi-step Creativity Tasks
Xiangyu Wang, Jin Wu, Xiaoyu Li, Chanjin Zheng and Yifeng Zhou 17272

IA-RAG: Interval-Algebra–Driven Temporal Reasoning for Dynamic Knowledge Retrieval
Xiaoman Wang, Yaoze Zhang, Wenzhuo Fan, Hongwei Zhang, Ding Wang, Guohang Yan, Song Mao, Botian Shi, Yunshi Lan and Pinlong Cai 17309

CHisAgent: A Multi-Agent Framework for Event Taxonomy Construction in Ancient Chinese Cultural Systems
Xuemei Tang, Chengxi Yan, Jinghang Gu and Chu-Ren Huang 17334

Grounding Structure in Language: Enriching Symbolic Representations for Molecular Reasoning
Zhengyuan Liu, Dan He, Nancy F. Chen and Wenya Wang 17362

A Four-Stage Decomposition of Word-Problem Solving and Mechanistic Fragility in LLM Math Reasoning
Zhongdi Qu and Carla P Gomes 17383

ReimburseBench: Benchmarking LMMs as Full-Workflow Auditors for Financial Internal Control
Libin Zhan, Wuyou Mao, Xiang Li, Enci Zhang, Qi Yang, Shujian Zhu, Yingtian Hu, Wentao Ye, Haobo Wang, Ziman Li, Jie Zhang, Wang Dong and Jun Chen 17403

PersonaTree: Structured Lifecycle Memory for Person Understanding in LLM Agents
Yubo Hou, Jingwei Song, Hongbo Zhang, Zhisheng Chen, Bang Xiao, Tao Wan and Zengchang Qin 17426

Can LLMs Access their World Knowledge for Event Prediction?
Zijun Wu, Zhitian Zhang, Hossein Hajimirsadeghi, Nikita Dvornik and Alexander Pashevich 17442

From AI Scientists to AI Scientific Communities: A Survey of Collective AutoResearch
Rui Zou, Tianyi Xu, Fan Zhang, Yao Zhang, Zhe Zhao, Xiaolongwang, Maolin Wang, Yingcheng Wu and Ye Wei 17458

TurnFlow: Benchmarking User-State Reasoning in Multi-Turn LLM Agents via Scene-Guided User Simulation
Zhixun Chen, Ming Li, Anjie Liu, Xiaoyuan Zhang, Quanyu Dai and Huangxun Chen 17473

ProtSynGen: Knowledge Graph-Guided Open-Ended Data Synthesis for Protein LLMs
Yifan Xu, Ying Lv, Gang Xu and Jianpeng Ma 17498

BanglaMemeX: Advancing Cultural Metaphoric Image Interpretation in Bangla with a Multimodal Explainable Dataset
Md. Sadman Sakib, Zisan Mahmud, Md. Fahim Arefin and Md Fahim 17525

SignFLIP: A Unified Model for Sign Language Translation and Generation via Stage-wise Alignment at Scale
Zhaoyi An, Sihan Tan, Youngbae Hwang, Kazuhiro Nakadai and Rei Kawakami 17562

Title It Right: Faithful and Attractive Paper Title Generation via Pareto Dominance Alignment
Ke Xiang, Yingchao Piao, Zexin Wang, Yufei Hou, Jianhui li, Jingjing LI, Zihan Liu, Hang Cui and Changhua Pei 17577

RecoverAgent: Intent-Preserving Recovery from Indirect Prompt Injection in Tool-using LLM Agents
Hanli Peng, Ziyao Liu, Yongsen Zheng and Kwok-Yan Lam 17596

Reinforcing Step-level Reasoning for Effective Self-Correction in LLMs
Duc Anh Vu, Nhat M. Hoang, Do Xuan Long, Cong-Duy T Nguyen, Ponhvoan Srey and Anh Tuan Luu 17613

OPD-Evolver: Cultivating Holistic Agent Evolver via On-Policy Distillation
Guibin Zhang, Xun Xu, Yanwei Yue, Zikun Su, Wangchunshu Zhou, Xiaobin Hu and Shuicheng Yan 17634

Clarify User Expertise: Towards Proactive Conversational Agents Tailoring Responses to User Proficiency
Zhihong Cao and Chen Huang 17653

Progress-SQL: Improving Reinforcement Learning for Text-to-SQL via Progressive Rewards
Shihao Zhang, Xiaoman Wang, Yuan Liu, Yunshi Lan and Weining Qian 17675

Dual-Group Exploration: Reinforcement Learning for Video Understanding in Multimodal Large Language Models
Qucheng Peng, Rui Hou, Jingyi Zhang, Shengjie Zhu, Yingfu Xu, Haotian Xu, Noranart Vesdapunt, Chen Chen and Ning Zhou 17696

TRONTOR: To RoPE or Not to RoPE for Long-Context Modeling
Youngjun Son, Jongyeon Park and Jaejin Lee 17720

Demystifying Entropy Conflict: Why Reasoning Compression Stalls in Large Reasoning Models
Hourun Zhu, Yang Gao, Wenlong Fei, Jiawei Li, Huashan Sun, Bowen Ren and Chengzhi Hu 17734

Write Policy Determines Source Retention in External FIFO Memory: A Delayed-Copy Diagnostic
Xiangyang Liu and ChunXia Yang 17748

When Does Generating More Help? Disentangling Fixed-Source Synthesis from Source Expansion in Synthetic Data Scaling
Xu Guo, Jian Tong, Zhihui Lu and Qipeng Guo 17767

Efficient Language-to-Vision Feature Injection for Referring Single-Object Tracking
Han Wang, Yuxuan Liu, Yuhan Sun, Jian Yang, Xiaotong Xu, Yixuan Lv, Zhuang Zhou and Shengyang Li 17784

CreaMem: A Scene-Aware Memory Architecture for Personalized Agents
Qixuan Sun, Yue Que, Bowei He, Jin Guo, Dihang Yang, Wenchang Situ and Chen Ma . . . 17800

ToolDF: Tool-Integrated Reasoning for Mixed-Authenticity Audio Deepfake Detection
Taewoo Kim, Young Han Lee, Nam IN Park and Chanwoo Kim 17814

CE-GAD: Curriculum-Efficient On-Policy Adversarial Distillation of Large Language Models
Wenhui Zhu, Hejian Sang, Xiwen Chen, Zhipeng Wang, Shayan Mohajer Hamidi, Han Yu, Jingjing Wang, Yuanda Xu, Xuanzhao Dong, Zhengze Zhou, Ran He, Jelena Markovic-Voronov, Kayhan Behdin, Sayan Ghosh and Alborz Geramifard 17827

Logic Tree: Topology-Aware Backtracking for DeepResearch Agents
Tianwen Lyu, Zhengke Gui, Peilong Zhao, Ling Zhong, Yiwen Zhang, Lei Liang, Mengshu Sun, Jun Zhou, Qiang Zhang and Huajun Chen 17837

KnowMT-Bench: Benchmarking Knowledge-Intensive Long-Form Question Answering in Multi-Turn Dialogues
Junhao Chen, Yu Huang, Siyuan LI, Rui Yao, Hanqian Li, Hanyu Zhang, Jungang Li, Jian Chen, Bowen Wang and Xuming Hu 17857

SCIT: Testing Causal Cache Carriers in Latent Chain-of-Thought Models
YI Ding, Lijun Huang and Menglin Yang 17905

CoT-Lock: Embedding Logic Fingerprints into the Chain-of-Thought of Large Reasoning Models
Jiaxin Hong, Hongyao Yu, Hao Fang, Kuofeng Gao, Yuxin Peng, Bin Chen, Shuhan Qi and Xuan Wang 17947

Beyond Redundancy: Information-Conserving Semantic Packaging for Vision-Language Models
Xin Xiao and Jiang Zhong 17970

Lexical Coupling in GUI Element Grounding: Sentence Embeddings Track Labels across Mobile and Web
Qijia Chen and Giulio Jacucci 17981

CAPO-SLT: Confidence-Aware Policy Optimization for Stable Sign Language Translation Generation
Ling Zhou, Yuhao Chen, Lin Cheng, Chengwen Yao, Donghui Sun and Xiaoxin Chen 17999

EntangleCodec: A Unified Discrete Audio Tokenizer via Semantic-Acoustic Entanglement
Hui Li, Yangfan Gao, Junlin Shang, Changhao Jiang, Tao Gui, Qi Zhang and Xuanjing Huang 18015

SCHEDBench: A Benchmark for Evaluating LLM Constraint Faithfulness in Natural-Language Combinatorial Scheduling
Shrenil Shaun Sharma and Avi Sharma 18032

Implicit Hierarchical Modeling for Cross-System Legal Case Retrieval
Zhengjia Zhong, Hongyi Lan, Yuhong Chen, Shengjia Hua, Kaiyan Liang, Xiaodong Li, Xianming Lin, Hui Li and Rongrong Ji 18053

MCTL: Meta-Cognitive Tool-Learning Framework for Multimodal Fake News Detection
Yue Han, WeiYao Zhang and ZhenZi Peng 18075

Um, Actually: Timely and Calibrated Fact-Checking with DIVER
Senyi Zhang, Tasnim Kabir and Jordan Boyd-Graber 18092

Neural Procedural Memory: Empowering LLM Agents with Implicit Activation Steering
Chengfeng Zhao, Yuqiao Tan, Shizhu He, Yequan Wang, Jun Zhao and Kang Liu 18107

In-Context Neurofeedback: Can LLMs Control Their Internal Representations through Privileged Access?
Koshiro Aoki, Ryota Takatsuki, Gouki Minegishi, Yusuke Haruki and Daisuke Kawahara . 18129

Scenario-based Probing and Steering Cultural Values in Large Language Models
Trung Duc Anh Dang, Tung Kieu and Sarah Masud 18153

EffectMem: Effectiveness-Driven Memory Reuse for Long-Horizon LLM Agents
Jingjing Yin, Lingbo Wang, Dongyan Cao and Yu Hong 18175

AdaSpec: Adaptive Routing by Ranking for Speculative Agents
Kangqi Ni, Wenyue Hua, Jinbiao Wei and Tianlong Chen 18196

From Correctness to Utility: Gain-Based Prefix Evaluation for LLM Reasoning
Yuhang Zhou, Yixin Cao and Guangnan Ye 18212

MERTA: Multi-Expert Retrieval for Token Allocation in Video Large Language Models
Bo Gao, Donghui Feng, Qi Wang, Qunshan Gu, Zhengxue Cheng and Li Song 18232

A Unified Framework to Elicit Structured Feedback for Interpretable Multi-Trait Essay Scoring
Shihang Yang, Sanwoo Lee, Ningning Zhao and Yunfang Wu 18247

GAGPO: Generalized Advantage Grouped Policy Optimization
Siyuan Zhu, Chao Yu, Rongxin Yang, Zongkai Liu, Jinjun Hu, Qiwen Chen and Yibo Zhang 18261

Estimating Uncertainty from Reasoning: A Large-Scale Study of Multi- and Crosslingual MCQA Performance in LLMs
Andrea Alfarano, Andrea Bacciu, Saab Mansour, Amin Mantrach and Marcello Federico . 18275

DisCon4Gen: A Semantic Alignment Framework Through Discriminative Conditioning for Generation
Caiwei Yang, Yanping Chen and Shuai Yu 18293

On the Learnability of Antidistillation Reasoning Traces
Hongyu Zhu, Sichu Liang, Wenwen Wang, Bofan Chen, Hanyi Wang and Shi-Lin Wang . . 18312

ElementCheck: Complexity-Aware Long-Form Text Factuality Evaluation via Sentence Elements
Xinming Wang, Haoran Du, Yi Chen, Xueqing Chen, Jian Xu, Hong-Ming Yang, Han Hu, Yulong Chen, Cheng-Lin Liu and Xu-Yao Zhang 18336

STQA: A Benchmark for Stock-Focused Tabular Question Answering over Historical and Forecasted Data
Baoxu An, Wenmian Yang, Zhensheng Wang and Weijia Jia 18364

Beyond Confidence: Test-Time Scaling for Multi-Turn Search Agents via Retrieval Grounding
Hyunho Kook, Junhyuk So, Tianyu Fu, Haizhong Zheng and Beidi Chen 18388

Hallucination as Exploit: Evidence-Carrying Multimodal Agents
Guijia Zhang, Hao Zheng and Harry Yang 18413

Personalizing How Recommendation Agents Reason: Self-Evolving Per-User Skills for Recommendation Agents
Zhen Tao, Riwei Lai, Chenyun Yu, Weixin Chen, Li Chen, Beibei Kong, Lei Cheng, Chengxiang Zhuo, Zang Li and Qingqiang Sun 18438

When Debate Messages Help or Hurt: Diagnosing Communication in Self-Distillation
Jiatao Yao and Bovin Tan 18461

SAMAC: Collaborating with Multi-Disciplinary Team for Adaptive Medical Decision-Making
Maosheng Li, Xiao Wu, Hu Wang, Imran Razzak and Yutong Xie 18474

DualResearch: Entropy-Gated Dual-Graph Retrieval for Answer Reconstruction in Deep Research Agents
Jinxin Shi, Zongsheng Cao, Runmin Ma, Yusong Hu, Jie Zhou, Yue Fan, Xin Li, Lei Bai, Liang He and Bo Zhang 18486

ESPO: Error-Structured Prompt Optimization via Diagnose, Diversify, and Stabilize
Lihao Liu, Peng Tang, Kunwar Yashraj Singh and Shabnam Ghadar 18502

VIBE: Video Instruction-aligned Background music gEneration
Aryan Vijay Bhosale, Vaibhavi Lokegaonkar, Vishnu Raj, Gouthaman KV, Sreyan Ghosh, Ramani Duraiswami, Lie Lu and Dinesh Manocha 18519

Out-of-Distribution Checkpoint Selection Has a Resolution Problem: Auditing Sparse Exact Match with Token Likelihood
Kunhee Ryu and Keeheon Lee 18535

Can LLMs Compress Wireless Signals? Physics-Informed Prompting for Training-Free Signal Compression
Yiwei Wang, Lixiang Lian and Sijie Ji 18553

Node-to-Neighborhood Semantic Consistency: Text-Topology Alignment for TAGs Anomaly Detection
Bochen Lin, Jianxiang Yu, Jiayi Wu, Lin QI, Lu Huang and Xiang Li 18574

MedReaMM: Evaluating Large Multimodal Models on Expert-Level Clinical Diagnostic Synthesis
Lai Wei, Yuchao Chen, Zhenbiao Cao, Xiaojin Zhang, Zhongyu Wei, Bangting Wang, Wei Chen and Xiang Bai 18590

CoCoA: Context-Conditional Cultural Alignment for Large Language Models
Kyungdon Lee, Wei Xu, Alan Ritter, Dong-Ho Lee and JinYeong Bak 18619

Think Before You Score: Pedagogical Chain-of-Thought Annotation and Reinforcement Learning for Low-Resource Multilingual Essay Evaluation
Wei Tian, Yuhao Zhou, Shenwei Li, Xiayu Sun and Man Lan 18641

Distilling LLM Reasoning into an Interpretable Policy Tree for Human-AI Collaboration
Beiwen Zhang, Yongheng Liang, Guowei Zou, Haitao Wang and Hejun Wu 18661

IndicDetect: Evaluating Cross-Lingual LLM-Generated Text Detection for Hindi, Telugu, and Tamil
Bhaskar Ganesh Devalla, Junchao Wu, Nilesh Dokuparthi, Greeshma Yaluru, Tatiana Muniz Rodriguez, Lidia S. Chao and Derek F. Wong 18683

Distilling Multimodal Evolution with Adaptive Anchors for Cloud-Edge Collaborative MLLMs
Chunyun Zhang, Qilong Song, Chaoran Cui, Shuai Gong, Fujun Yang, Wenna Wang and Peiguang Lin 18732

Doc-CoB: Enhancing Document Understanding with Visual Chain-of-Boxes Reasoning
Ye Mo, Kai Ye, Xianwei Mao, Zirui Shao, Gang Huang, Bo Zhang, Hangdi Xing, Kehan Chen, Zhou Huan, Zixu Yan, Jiajun Bu and Sheng Zhou 18744

FORGE: Fine-grained multimodal evaluation for manufacturing quality assurance
Xiangru Jian, Hao Xu, Wei Pang, Xinjian Zhao, Chengyu Tao, Qixin Zhang, Xikun Zhang, Chao Zhang, Guanzhi Deng, Jingxi Xue, Juan Du, Tianshu Yu, Garth Tarr, Linqi Song, Qiuzhuang Sun and Dacheng Tao 18762

Skill-as-Pseudocode: Refactoring Skill Libraries to Pseudocode for LLM Agents
Xinze Li, Yuhang Zang, Yixin Cao and Aixin Sun 18840

AEGIS-D: Multi-Agent Debate for High-Precision Verification of Hard Competition Mathematics
Jiayang Sun, Jiawei Xu, Maxm Pan, Zerui Cheng and Pramod Viswanath 18861

STARS: Skill Triage via Activation Risk Scoring for LLM Agents
Guijia Zhang, Shu Yang, Xilin Gong and Di Wang 18874

CoopAgent: Training Multi-Agent Cooperation from Single-Agent Execution Trajectories
Zhengzhe Pan, Linchao Zhu and Yi Yang 18891

Evaluating LLM Agents Beyond Local Edit Validity
Kunhee Ryu, Chi-Guhn Lee and Keeheon Lee 18910

PersonaEdit: Representative Sample Selection for Personalized Model Editing
You-Mei Huang, Chung-Chi Chen and An-Zi Yen 18945

AMU-GEC: A Polish Native Speaker Grammatical Error Correction Benchmark
Ryszard Staruch, Jarosław Liberek, Krzysztof Jassem, Filip Graliński, Michalina Klimczak and Sonia Ołubczyńska 18961

The Struggle Between Continuation and Refusal: A Mechanistic Analysis of the Continuation-Triggered Jailbreak in LLMs
Yonghong Deng, Zhen Yang, Ping Jian, Xinyue Zhang, Zhongbin Guo, Chengzhi Li and Junxi Yin 18972

VALCON: Benchmarking and Enhancing Value Consistency in Large Language Models via Dual-Role Variation
Junru Wu, Jingting Zheng, Yuqi Ren and Deyi Xiong 18988

MemGovern: Enhancing Code Agents through Learning from Governed Human Experiences
Qihao Wang, Ziming Cheng, Shuo Zhang, Fan Liu, Rui Xu, Heng Lian, Kunyi Wang, Xiaomin Yu, Jianghao Yin, Sen Hu, Yue Hu, Shaolei Zhang, Yanbing Liu, Ronghao Chen and Huacan Wang 19012

TokenPilot: Cache-Efficient Context Management for LLM Agents
Buqiang Xu, Zirui Xue, Dianmou Chen, Chenyang Fu, Chiyu Wu, Caiying Huang, Chen Jiang, Jizhan Fang, Xinle Deng, Yijun Chen, Yunzhi Yao, Xuehai Wang, Jin Shang, Huajun Chen and Ningyu Zhang 19034

Romanized Arabic Across Dialects: Views, Usage Patterns, and Linguistic Variation
Amr Keleg, Ahmed Amine Ben Abdallah, Taha Yassine, Chadi Helwe, Imane Guellil and Nedjma Ousidhoum 19050

HeTGB: A Comprehensive Benchmark for Heterophilic Text-Attributed Graphs
Shujie Li, Yuxia Wu, Yuan Fang and Chuan Shi 19080

Alignment by Stereotyping: How LLMs Sacrifice Individual Distinctiveness for Cultural Adaptation
Qishuai Zhong, Zongmin Li, Siqi Fan and Aixin Sun 19097

EM²Mem: Event-Centric Multimodal Memory for Large Language Models
Yijun Chen, Yaqi Zheng, Yanya Li, Boyi Xiao, Buqiang Xu, Shuofei Qiao, Jizhan Fang, Xinle Deng, Yunzhi Yao, Xuehai Wang, Liuxin Zhang, Hui Li, Huajun Chen and Shumin Deng 19117

AIMER: Calibration-Free Task-Agnostic MoE Expert Pruning
Zongfang Liu, Guangyi Chen, Yifan Shen, Shengkun Tang, Huan Wang and Xin Yuan 19145

EMemBench: Interactive Benchmarking of Episodic Memory for VLM Agents
Xinze Li, Ziyue Zhu, Siyuan Liu, Yubo Ma, Yuhang Zang, Yixin Cao and Aixin Sun 19163

LiST: Local-Simplex Test-Time LoRA Fusion
Yihua Shao, Jia Li, Siyu Chen, Xinyu Luo, Yang Liu, Kecheng Chen, Xinwei Long, Lingyu Zhu, Fanhu Zeng, Maolin Wang, Ziyang Yan, Jingcai Guo, Hao Tang, Nicu Sebe and Zhenyi Wang . . 19193

ChapterRL: Chapter-Aware Reward Framework for Novel Generation Reinforcement Learning
Bingqian Sun, Jingchao Yang, Xiangyu Wu, Zhang Yunfei, Tingting Gao and Han Li 19206

Activation-Space Anchored Access Control for Multi-Class Permission Reasoning in Large Language Models
Zhaopeng Zhang, Pengcheng Sun, Lan Zhang, Chen Tang, Jiewei Lai, Xirong Zhuang, Yunhao Wang, Hui Jin and Li Liu 19224

Improving Low-resource Grammatical Error Correction Via Automatic Linguistic Neurons Probing and Optimizing
Zihao Yu, Ying Li, Zhengtao Yu, Shengxiang Gao, Cunli Mao, Yuxin Huang, Hua Lai and Wenjun Wang 19241

JudgePanel: A Compact Judge with Panel Deliberation via Adaptive Multi-Reward Reinforcement Learning
Yiyue Qian, Shinan Zhang, Huan Song and Hannah R Marlowe 19253

AFCS: Atomic-Fact-Conditioned Synthesis for Factual Knowledge Pretraining
Yunhao Wang, Feng Zhang, TingHao YU, Xing W and Jiang Zhou 19268

IRIS: Iterative Refinement of Internal States for Safer Language Models
Beize Zhang, Yuchen Feng, Peize Li, Zhitong Zhu, Peng Fu, Mu Weimin, Zheng Lin and Weiping Wang 19288

Probing Benchmark-Specific Regularities in Knowledge Gating via Lightweight Feature Analysis
Zilong Cheng, Yubo Chen, Jingbo Sun, Ge Gao, Bo Chen and Xiaobing Zhao 19303

CrossAlpha: An Annual-Report Benchmark for Cross-Market Factor Research with LLM Agents
Qian Wang, Tong Zhongyi, Nuo Chen, Zhaomin Wu and Bingsheng He 19312

AEMAgent: Cognition-Inspired Active Memory Formation for Very-Long Video Understanding
Xian Zhao, Jiaming Zhang, Haonan Chu, Weihao Bo, Xu Chen, Delai Qiu, Yining Wang, Yi Zhang, Shengping Liu, Jian Yu and Jitao Sang 19342

Talking to Itself While Coding: What Makes Comments Help Code Generation?
Dangfeng Pan, Zhensu Sun, Cenyuan Zhang, David Lo and Xiaoning Du 19356

Skill0.5: Joint Skill Internalization and Utilization for Out-of-Distribution Generalization in Agentic Reinforcement Learning
Jiapeng Zhu, Jianxiang Yu, Yibo Zhao, Chengcheng Han, Xierui Song, Xunliang Cai, Xiang Li and Weining Qian 19381

Learning New Facts with QLoRA: An Acquisition-Retention Frontier
Estelle Zheng, Sébastien Warichet, Emmanuel Helbert and Christophe Cerisara 19397

One prompt is not enough: Instruction Sensitivity Undermines Embedding Model Evaluation
Yevhen Kostiuk and Kenneth Enevoldsen 19408

SkillMol: Skill-Conditioned Action Editing for Molecular Optimization
Jiajun Yu, Kun Li, Zhihao Wu, Weifan Mao, Yizhen Zheng and Haishuai Wang 19428

Expected Contribution Matching: Optimal Value Merging for KV Cache Compression
Pengzhan Li, Yixuan Wang and Wanxiang Che 19446

Enhancing Event Candidate Acquisition for Event Linking
Ziyang Zhang, Yinan Liu, Boyi Xue, Yingxuan Huang, Bin Wang and Xiaochun Yang 19459

DeALOG: Decentralized Multi-Agents Log-Mediated Reasoning Framework
Abhijit Chakraborty, Ashish Raj Shekhar, Shiven Agarwal and Vivek Gupta 19470

Building Long-Term LLM Memory with Neo-Davidsonian Discourse Units
Sizhe Zhou, Yanzhen Shen, Siru Ouyang, Yuyang Bai, Yiheng Shu, Yu Zhang and Jiawei Han 19494

AQS: MoCA-Guided Adaptive Questioning with Learned Stopping for Efficient Cognitive Impairment Screening
Mingyu Li, Shuqiong Wu and Zhaojie Luo 19532

GRAIL: Gradient-Reward Aligned Initialization for LoRA-based GRPO
Liwen Shi, Zijie Zeng, Bowen Wang, Yunji, Zeyu Chen, Xinyu Shao, Xinzheng Sun, Xuyang Xiong, Haoyang Wu and Sheng Zhang 19549

LegalOne: A Family of Foundation Models for Reliable Legal Reasoning
Haitao Li, Yifan Chen, Shuo Miao, Qian Dong, Jia Chen, Yiran HU, Junjie Chen, Minghao Qin, Yueyue Wu, Yujia Zhou, Qingyao Ai and Yiqun Liu 19566

The Shape of Power: A Multilingual Framework for Social Power Reasoning in Dialogues
Farah Atif, Sougata Saha and Monojit Choudhury 19589

Think Before You Prune: Selective Self-Generated Calibration for Pruning Large Reasoning Models
Yang Xiang, Yixin Ji, Juntao Li and Min Zhang 19605

ProMediConv: Benchmarking Proactive Conversational Agents in Legal Dispute Mediation
Zesheng Wei, Mengfan Li, Wenhao Liu, Yixin Zhang, Zilei Wang and Yang Deng 19618

Unlocking the Unsolvable: Teacher-Guided Curriculum for Data-Efficient RLVR
Yukang Zhu and Zhen Han 19639

EnterpriseBench: Benchmarking LLM Agents on Enterprise-Level Strategic Reasoning and Decision-Making
Min Yang, Yichen Pan, Jinghua Piao, Dandan Song, Yongshun Gong and Yong Li 19659

How You Ask Shapes What You Get: A Theory-Seeded Measurement of Articulation in Advice-Seeking LLM Conversations
Juneha Baek, Suhyeon Lee and Donghyuk Shin 19683

Skip-Connected Rollouts for Mid-Trajectory Rewards in GRPO
Fengwei Teng, Jinyi Bai, Xinhao Yao, Demi Wang, Jiahao Zhao and Zhijiang Guo 19710

SYMMETRA : Semantic prompt sanitization for end-to-end utility-preserving privacy protection in LLM interactions
Seongsu Moon, Juyeon Jeong and Taejune Park 19729

Test-Time Training with Next-Token Prediction
Xuan Ouyang, Zefan Cai and Junjie Hu 19750

TRIO: Task Refinement with Individual Optimization for Retrieval-Augmented Generation
Yingtao Ren, Xiao Luo, Yiwei Fu, Jinzhao Zhou, Yu-Cheng Chang and Chin-teng Lin 19765

The Target Is a Predicate: Predicate-State Grounding for Semantic-Level Video Reasoning Segmentation
Yuhuan Hong, Yiming Ren, Zhiqiang Lin, Feize Wu, Liu Xingyun, Yitong Xie, Lei Zhou, Lisheng Wu, Junjie Wang and Yujiu Yang 19783

LongMem-RAT: Long-Term Memory for Cross-Session Red Teaming
Nurit Cohen Inger, Barak Haziz, Seffi Cohen, Yonatan Diga, Barak Chizi, Lior Rokach and Bracha Shapira 19798

Triplet-Block Diffusion RWKV
Ke Lin, Yiyang Luo, Zhaolong Su, Yunya Song and Anyi Rao 19816

How to Score Experts for One-Shot MoE Expert Pruning: A Systematic Formulation and Selection Principle
Zongfang Liu, Jinghui Zhang, Zijian Ma, Guangyi Chen and Xin Yuan 19830

Rank4Gen: RAG-Preference-Aligned Document Set Selection and Ranking
Yongqi Fan, Yuxiang Chu, Zhentao Xia, Xiaoyang Chen, Jie Liu, Haijin Liang, Jin Ma, Ben He, Yingfei Sun, Jie Zhai, Dezhi Ye and Tong Ruan 19847

K-BrowseComp: A Web Browsing Agent Benchmark Grounded in Korean Contexts
Nahyun Lee, Dongkeun Yoon, Guijin Son, Geewook Kim, Dayoon Ko, Jeonghun Park, Haneul Yoo, Jaewon Cho, Junghun Park, Changyoon Lee, Kyochul Jang, Jaeyeon Kim, Eunsu Kim and Seungone Kim 19876

TypoCD: Training-Free Contrastive Decoding for Typographic Defense in Large Vision-Language Models
Xia Sihan, Jing Li, Junhao Dong, Ivor Tsang and Yew-Soon Ong 19905

AdaptInfer: Adaptive Token Pruning for Vision Language Model Inference via Dynamical Text Guidance
Weichen Zhang, Zhui Zhu, Ningbo Li, Shilong Tao, Hongzi Zhu, Jingao Xu, Kebin Liu and Yunhao Liu 19920

EgoErrorVQA: Assess Egocentric Comprehension Capabilities through Procedural Errors for Ego-Agentic AI
Junlong li, Junxi LI, Jianjun Gao, Chen Cai, Lap-Pui Chau and Yi Wang 19939

MemPulse: Memory Pulse-Driven Adaptive Long-Context Reasoning with Fact-based Hypergraph
Wenbiao Tao, Yuanfu Wang, Yunshi Lan and Chao Yang 19960

Function-Level Execution Feedback for Code Preference Optimization
Idris Nechnech, Sehwan Kim, Jimin Seo, Yeongoon Kim, Minhae Oh, Sangwoo Hong and Jungwoo Lee 19976

MI-Distillation: Selecting from Model-Interpolated Instruct-Reasoning Data Spectrum for Chain-of-Thought Distillation
Yangsong Lan, Renkai Hu, Hong-Kai Zheng, Bo Zhang, Renzhi Wang, Hongliang Dai and Piji Li 19996

Improving Synthetic Data Generation for Argument Mining via Adversarial Reinforcement Learning
Zhijun Zhang, Qianlong Wang, Keyang Ding, Genan Dai, Bowen Zhang, Bin Liang, Ruifeng Xu and Yongsheng Liang 20017

SCOPE-SQL: Structured Semantic Credit for Reinforcement Learning of Interactive SQL Repair Agent
Hong Qi, Junyi Zhao, Boyan Li, Juyong Jiang and Yuyu Luo 20037

Forget-to-Focus: Targeted Unlearning for Domain Specialization in LLMs
Chaitali Bhattacharyya, Shinhyoung Jang, Seohyun Kim, Junyoung Lee and Yeseong Kim 20059

Beyond Factual QA: Mentorship-Oriented Question Answering over Long-Form Multilingual Content
Parth Bhalerao, Ruiwen Guan, Diola Dsouza and Oana Ignat 20086

Are LLMs Ready to Assist Physicians? PhysAssistBench for Interactive Doctor-Patient-EHR Assistance
Tianming Du, Peijie Yu, Sihan Shang, Danli Shi, My Linh Nguyen, Shengbo Gao, Guangyuan Li, Yinghong Yu, Yan Jiang, Qianlong Zhao, Behzad Bozorgtabar, Shaoxiong Ji, Jiazhen Pan, Daniel Rueckert and Jiancheng Yang 20108

SnapBench: Benchmarking Snap-and-Ask Multimodal Retrieval for Mobile Interactions
Zirong Chen, Fuda Ye, Kuan Zhang, Enjun Du, Junfu Pu, Xinlei Wang, Xinyu Zuo, Lisheng Duan, Jin Ma and Yongqi Zhang 20143

Can Agents Read the Room? Benchmarking Visual Social Intelligence in Multimodal Simulation
Shijun Wan, Xuehai Wu, Jiwen Zhang, Siyuan Wang and Zhongyu Wei 20179

TACME: Task-Aware Cross-Modal Experts for Disjoint-Corpus Multimodal Behavioral Understanding
Elena Ryumina, Alexandr Axyonov and Dmitry Ryumin 20210

CogManip: Benchmarking Manipulative Behavior in Multi-Turn Interactions with Large Language Models
Zeyang Yue, Chenfei Yan, Feifei Zhao, Haibo Tong, Mengwen Xu, Xiaozhen Wang, Erliang Lin and Yi Zeng 20235

Robust Text Watermarking for Large Language Models via Dual Semantic Embeddings
Jonas Schäfer, Cezary Pilaszewicz and Gerhard Wunder 20266

AdaR: A Framework for Equipping LLMs with Adaptive Reasoning
Zhejian Lai, Xiang Geng, Zhijun Wang, Yang Bai, Jiahuan Li, Rongxiang Weng, Jingang Wang, Xuezhi Cao, Xunliang Cai and Shujian Huang 20288

DeepWeaver: Bridging the Evidence Synthesis Gap in Open-Ended Question Answering
Xujia Wang, Yizhe Zhang, Bin Xu, Lei Hou and Juanzi Li 20307

One-Stage Multi-Task Instruction-Guided 3D Spatial Audio Editing
Ke Lei, Chenyuhao Wen, Yu Zhang, Wenxiang Guo, Changhao Pan, Sashuai Zhou, Yongshi Li, Ruiqi Li, Ruofan Hu, Haorui Xu, Xiang Yin and Zhou Zhao 20355

Internalizing Meta-Experience into Memory for Guided Reinforcement Learning in Large Language Models
Shiting Huang, Zecheng Li, Yu Zeng, Qingnan Ren, Zhen Fang, Qisheng Su, Kou Shi, Ziao Zhang, Lin Chen, Zehui Chen and Feng Zhao 20370

Diagnosing and Repairing Local Reasoning Preference Collapse in Medical Chain-of-Thought Distillation
RuiQing Wang, Xiaoyu Tan, Xiyao Wang, Jue Chen and Xihe Qiu 20390

Learning to Recover Bridge Path for SQL Schema Linking via Tentative Answer Profiling
Wenzhong Qiu, Shixin Liu and Yu Hong 20413

Data Augmentation via Relation-Variant Discourse Recasting for Implicit Discourse Relation Recognition
Yikang Yang, Weihao Sun, Yu Hong, Zhipang Wang and Guodong Zhou 20430

Exploring Information Seeking Agent Consolidation
Guochen Yan, Jialong Wu, Zhengwei Tao, Bo Li, Qintong Zhang, Jiahao Xu, Haitao Mi, Yuejian Fang, Qingni Shen, Wentao Zhang and Zhonghai Wu 20445

From Monolithic Prompts to Scaffolded Execution for Memory Agents with Small Language Models
Seunghan Yang, Jihwan Bang, Sungha Choi and Juntae Lee 20471

Light or Full Verb? A Minimal-Pair Dataset for Probing Phraseological Competence in Language Models
Francesca Franzon, Nicolás Rosàs Gómez and Leo Wanner 20499

KARMA: Knowledge graph-based Automated Reasoning Materialization and Alignment
Jinkyeong Choi, Chaebin Jeong and Donghyeon Park 20513

Can Agentic Trading Systems Pay for Their Own Intelligence?
Qiqi Duan, Changlun Li, Chen Wang, Fan Zhang, Mengxiang Wang, Dayi Miao, Peixian Ma, Jiangpeng Yan, Liyuan Chen, Shuoling Liu, Preslav Nakov, Yuyu Luo and Nan Tang 20532

From Homogeneity to Diversity: Self-Refined GRPO for Controllable Creative Writing
Minna Peng, Wei Tan, Zeyu Wang, Yongquan Hu, Ruizhi Huang and Yang Jie 20555

ASE-MCTS: Asynchronous Self-Evaluated Monte Carlo Tree Search for Efficient LLM Reasoning
Yang Wu, Yuyang Liu, Zhicheng Zhang and Xiaolong Zheng 20572

AutoCRAT: Within-trajectory Joint Control of Stochasticity and Compute for LLM Reasoning
Hanjun Luo, Qiushi Liu, Jingya Zhang, Haihong Pang, Jiaheng Wen, Yifei. Ma, Yu Yao, Chengxi Zhang, Hanrong Zhang, Yankai Chen and Hanan Salam 20592

MoE Sparsity Begets Monosemanticity: Insights into Understanding and Fine-Tuning
Mingxue Tian, Qi Zhang, Xiaohan Wang, Jiajun Chai, Guojun Yin, Wei Lin, Cailian Chen, Qimin Xu, Siyu Wang and Yisen Wang 20613

Isolation as a First-Class Principle for LLM-Agent System Safety: Concepts, Taxonomy, Challenges and Future Directions
Huihao Jing, Wenbin Hu, Shaojin Chen, Haochen Shi, Sirui Zhang, Hanyu Yang, Changxuan Fan, Zhongwei Xie, Hongyu Luo, Wun Yu Chan, Wei Fan, Haoran Li and Yangqiu Song 20631

Formally Verified AI Co-Reasoning: An Education Case Study
Arnav Mehta, Chenjun Guo, Thomas Lu, Niels Voss, Manooshree Patel, Rayna Bhattacharyya, Peter Donovan and Gireeja Ranade 20651

MissMAC-Bench: Building Solid Benchmark for Missing Modality Issue in Robust Multimodal Affective Computing
Ronghao Lin, Honghao Lu, Ruixing Wu, Aolin Xiong, Qinggong Chu, Qiaolin He, Sijie Mai and Haifeng Hu 20669

MACRO-LLM: LLM-Empowered Multi-Agent Collaborative Reasoning under Spatiotemporal Partial Observability
Handi Chen, Running Zhao, Xiuzhe Wu, Zhanfeng Xu and Edith Cheuk-Han Ngai 20685

Multimodal Reward Hacking in Reinforcement Learning
Jiayu Yao, Yiwei Wang, Anmeng Zhang, Zhe Sun, Songsong Wang, Lingrui Mei, Yuyao Ge and Shenghua Liu 20718

Vital Trace: Protocol-Constrained Patient-State Reasoning for Longitudinal Clinical Trajectories
Zhan Qu and Michael Färber 20734

AsyncTool: Evaluating the Asynchronous Function Calling Capability under Multi-Task Scenarios
Kou Shi, Ziao Zhang, Shiting Huang, Avery Nie, Zhen Fang, Qiuchen Wang, Lin Chen, Huaian Chen, Zehui Chen and Feng Zhao 20752

In-Context Compilation: Encoding Context-Induced Behavioral Shift into Model Weights
Michael Merzin, Yehonatan Elisha, Oren Barkan, Yonatan Toib and Noam Koenigstein . . . 20777

Zipping the Thought: When and How Compressed Reasoning Data Works in LLM Post-Training
Kohsei Matsutani, Gouki Minegishi, Takeshi Kojima, Yusuke Iwasawa and Yutaka Matsuo 20797

Disentangling Form and World Knowledge in LLM Interpretation: Evidence from Quantifier Scope Disambiguation
Jakub Kosterna, Justyna Grudzińska, Maciej Miecznikowski, Wojciech Borysewicz, Julia Poteralska, Kacper Rutkowski and Jan Henryk Kwapisz 20823

Adapting Dynamic Sampling to Fine-Grained Rewards for Tool Learning with Curriculum Learning
Zihao Feng, Xiaoxue Wang, Bowen Wu, Zhen Xu, Hailong Cao, Tiejun Zhao and Baoxun Wang 20834

Beyond Polarization: The Generative Constraint of Chain-of-Thought in Pointwise Reranking
Xiaoyang Chen, Jie Liu, Haijin Liang, Haibo Shi, Jin Ma, Ben He, Yingfei Sun and Dezhi Ye 20854

CASCADE Against Jailbreaks: Combination Across Stages with Controlled Attack-Defense Evaluation
Jiale Luo and Han Liang Wee Eric 20866

PruneTIR: Inference-Time Tool Call Pruning for Effective yet Efficient Tool-Integrated Reasoning
Luan Zhang, Dandan Song, Zhijing Wu, Zhengyu Chen, Chen Zhang, Yuhang Tian, Huipeng Ma, Chenhao Li, Changzhi Zhou, Xudong Li, Ruochen Zhou and Shuhao Zhang 20889

FedSubMuon: Communication-Efficient Federated LLM Fine-Tuning via Structured Subspace Muon
Shaolong Chen, Youming Tao, Shuzhen Chen, Falko Dressler, Qingqing Ye and Di Wang . 20908

SKILLREVISE: Improving LLM-Authored Agent Skills via Trace-Conditioned Skill Revision
Yuxuan Liu, Zhaochen Su, Lingyun Xie, Yuhao Zhang, Qing Zong, Jiahe Guo, Zhongwei Xie, Yiyan Ji, Yauwai Yim, Hongyu Luo, Xiyu Ren, Ruan Chenyu, Haoran Li and Yangqiu Song . . . 20923

StreamMemBench: Streaming Evaluation of Agent Memory for Future-Oriented Assistance
Guanming Liu, Yuqi Ren, Hansu Gu, Peng Zhang, Weihang Wang, Jiahao Liu, Ning Gu and Tun Lu 20947

SAC-Copula: Quality-Preserving Watermarking for Diffusion Language Models via Smooth Correlated Gumbel Fields
Baixin Li and Haiyun He 20961

Self-Evolving Dual-Role Deliberation for Medication Recommendation from Clinical Records
Jiyue Jiang, Yu Lei, Yitao Xu, Zikang Wang, Zhenyu Wang, Licheng Zong, Han Huang, Ziyi Li, Zhiyu Huo, Yuhan Chen, Jiayang Chen, Xiaodan Fan and Yu Li 20985

Routing by Reasoning Need: Trajectory-Aware Decoding Control for Diffusion Vision-Language Models
Yixiang Liu, Zhongxing Xu, Zhonghua Wang and Xiaoying Tang 21005

Planning Before Reasoning: A Vision-Language Dataset for Structured Stepwise Reasoning in Document Understanding
Dongjae Shin, Hangyeol Yoo, Yerin Nam, Hoyeon Shin, SeungWoo Song, Seohyun Song, Donghun Yang and KyungTae Lim 21022

LoSATok: Low-Dimensional Semantic-Acoustic Tokenizer for Cross-Domain Audio Understanding and Generation
Zhisheng Zhang, Xiang Li, Yixuan Zhou, Jing Peng, Guoyang Zeng and Zhiyong Wu 21062

Mind the Implicit Gap: Quantifying Attribute Inference Leakage in Privacy-Conscious Delegation
Nicole Ren and James Teo 21080

A Systematic Comparison of Multilingual Interpretability Methods Reveals Anisotropy-Driven Failures
Oskar Holmström, Marcel Bollmann and Marco Kuhlmann 21115

Selective Agent Guidance via Entropy: Learning Autonomous Policies from Imperfect VLM Teachers
Giovanni Bonetta, Matteo Merler, Davide Zago, Rossella Cancelliere and Bernardo Magnini 21165

SAPA: a Situation-Augmented Personalized web Agent
Qianyue Li, Bin Ji, Xiaodong Liu, Shasha Li, Xiaopeng Li, Yueyu Wang, Xinran Hong, Ma Jun and Jie Yu 21192

DataChef: Cooking Up Optimal Data Recipes for LLM Adaptation via Reinforcement Learning
Yicheng Chen, Ma Zerun, Xie Xinchen, Yining Li and Kai Chen 21208

Multi-Step Knowledge Interaction Analysis via Rank-2 Subspace Disentanglement
Sekh Mainul Islam, Pepa Atanasova and Isabelle Augenstein 21231

BiasGym: A Simple and Generalizable Framework for Analyzing and Removing Biases through Injection
Sekh Mainul Islam, Nadav Borenstein, Siddhesh Milind Pawar, Haeun Yu, Arnav Arora and Isabelle Augenstein 21249

LoopViT: Scaling Visual ARC with Looped Transformers
Wen-Jie Shu, Xuerui Qiu, Rui-Jie Zhu, Harold Haodong Chen, Yexin Liu and Harry Yang 21271

X-GS: An Extensible Framework for Perceiving and Thinking with 3D Gaussian Splatting
Yueen Ma, Zenglin Xu and Irwin King 21281

TreePS-RAG: Tree-based Process Supervision for Reinforcement Learning in Agentic RAG
Tianhua Zhang, Kun Li, Junan Li, Yunxiang Li, Hongyin Luo, Xixin Wu, James R. Glass and Helen M. Meng 21297

LaME: Learning to Think in Latent Space for Multimodal Embedding via Information Bottleneck
Peixi Wu, Biao Yang, Feipeng Ma, Bosong Chai, Bo Lin, Wei Yuan, Fan Yang, Tingting Gao, Hebei Li and Xiaoyan Sun 21317

The Unsampled Truth: Quantifying Prompt Artifacts in LM Psychometrics
Nils Schwager, Christoph Hau, Simon Münker and Achim Rettinger 21330

Syntax-Aware Detection of Training Data Contamination in Code LLMs
Jiazheng Zhang, Cheng Liu, Long Cheng, Xiaowei Li and Huawei Li 21342

From Route to Survey: Probing Spatial Perspective Transformation in Large Language Models
Haotong Zhang 21368

Better World Models Can Lead to Better Post-Training Performance
Prakhar Gupta, Henry Conklin, Sarah-Jane Leslie and Andrew Lee 21387

SAP: State-Guided Data Synthesis with Argument Provenance for Multi-Turn Tool Use
Zichen Tian, Jinpeng Chen, Cheng Gong, Suiyun Zhang and Rui Liu 21400

Adaptive Hierarchical Representation Alliance for Multimodal Learning
Chunlei Meng, Pengbin Feng, Jacqueline J. Pang, Chih-Ting Liao, Rong Fu, Zhaolu Kang, Zhongxue Gan and Chun Ouyang 21423

SALA: Semantic-Aware Logical Alignment for Complex Reasoning in In-Context Learning
Zhao Ji, Wenqing Chen, Zhixuan Chu, Jianxing Yu, Jingping Liu, Shanhe Zhao and Zibin Zheng 21440

IFMTBench: A Comprehensive Benchmark for Multilingual Translation Instruction Following
Mingrui Sun, Mao Zheng, Zheng Li and Mingyang Song 21457

ViPlan: A Benchmark for Visual Planning with Symbolic Predicates and Vision-Language Models
Matteo Merler, Nicola Dainese, Minttu Alakuijala, Giovanni Bonetta, Pietro Ferrazzi, Yu Tian, Bernardo Magnini and Pekka Marttinen 21469

Breaking the Cloak! Unveiling Chinese Cloaked Toxicity with Homophone Graph and Toxic Lexicon
Xuchen Ma, Jianxiang Yu, Wenming Shao, Bo Pang and Xiang Li 21517

To Each Language Its Tokenizer: Modular Tokenizers for Efficient Multilingual LLMs
Franck Signe Talla, Hippolyte Pilchen, François Yvon and Edouard Grave 21534

Phonemizing User-Generated Text: A Benchmark, Taxonomy, and Compositional Approach
MinJu Jeon, Younghan Park, Han Sung Park, Jong-Hwan Kim, Dong-Jin Kim and Hoyeon Lee 21564

Can Released LLM Vocabularies Support Token-Level Estimation of Hidden Corpora?
Qingjie Zhang, Xingzhang Ren, Zixuan Chen, Jinfeng Li, YueFeng Chen, Yitong Yang, Hui Xue, Dayiheng Liu and Han Qiu 21582

SkillDAG: Self-Evolving Typed Skill Graphs for LLM Skill Selection at Scale
Tong Bai, Zhenglin Wan, Pengfei Zhou, Xingrui Yu, Yang You and Ivor Tsang 21597

Why Do LLM Agents Fail in Exploring New Environments? A World-Modeling Perspective
Shiqi Chen, Tongyao Zhu, Zian Wang, Jinghan Zhang, Kangrui Wang, Ruochen Zhou, Siyang Gao, Teng Xiao, Yee Whye Teh, Junxian He and Manling Li 21612

Interference-Aware Composition of Steering Vectors
Giwon Hong, Gorjan Radevski, Goran Glavaš, Kiril Gashteovski, Edoardo Ponti, Carolin Lawrence and Pasquale Minervini 21627

BASS: Budget-Aware Structured Search for Text-to-SQL Generation
Bin Zheng, Zheng Chen, Kefan Chen, Lulu Cui, Hao Tang, Zijun Xu and Chen Anlong . . . 21659

DLSA: Dual-Level Semantic Alignment with Adaptive Multi-Scale Modeling for Referring Remote Sensing Image Segmentation
Wei Qu, Wei Wei, Shihao Zou and Zian Cao 21687

DSFlow: Evolutionary Workflow Optimization for Generalizable LLM-Based Data Science Automation
Fan Liu, Tianyu Pang, Chao Du and Hao Liu 21705

WaymoQA: A Multi-View Visual Question Answering Dataset for Safety-Critical Reasoning in Autonomous Driving
Seungjun Yu, Seonho Lee, Namho Kim, Jaeyo Shin, Junsung Park, Wonjeong Ryu, Raehyuk Jung and Hyunjung Shim 21721

Towards Reliable, Generalizable, and Specific In-Context Knowledge Editing via Multi-Objective Reinforcement Learning
Xuzhong Wang, Maiqi Jiang, Tejal Nair, Girija Bhusal, Yanfu Zhang and Haipeng Chen . . 21745

ReVEL: Multi-Turn Reflective LLM-Guided Heuristic Evolution via Structured Performance Feedback
Van Duc Cuong, Nguyen Dinh Tuan Minh, Vu Duc Tam, Tung D. Vu, Son Van Nguyen, Nguyen Thi Hanh and Huynh Thi Thanh Binh 21762

Same Symptoms, Different Severity?: Investigating Gendered Cue Effects on LLMs in Depression Severity Assessment
Migyeong Kang, Sihyeon Yang, Minseok Kim, Jisun An, Haewoon Kwak and Jinyoung Han 21783

Toward Detecting Modern Social Media Agents: Dataset, Experimental Analysis, and an Agent-Based Solution
Yi Han, Lizi Liao, Haiqi Lu, Yuanxing Liu, Weinan Zhang and Ting Liu 21801

SEC: Selective Embedding Correction for Cross-Lingual Transfer in Low-Resource Languages
Mengting Hu, Wenjie Xue, Peiyu Xu, Xuan Li, Junpeng Huang, Zhiyuan Guang, Xiaoru Zhang, Hang Gao, Sicheng Wei, Bingyang Fan and Tiegang Gao 21821

The Physicist, the Coder, and the Critic: Agentic Photonic Crystal Fiber Inverse Design via Verbal Gradient Descent
Shengchao Chen and Ting Shu 21837

Do AI Reviewers Converge? Diversity Collapse in Model-Generated Peer Review
Xuzhe Hou, Shengxin Zhang, Riley Stacy, Emily Black and Varun Chandrasekaran 21856

Controlled Self-Evolution for Algorithmic Code Optimization
Tu Hu, Ronghao Chen, Shuo Zhang, Rui Xu, Jianghao Yin, Mou Xiao Feng, Jingping Liu, Wenqi Jiang, Shaolei Zhang, Yuqi Fang, Sen Hu, Yi Xu and Huacan Wang 21886

Improving Allocentric Spatial Reasoning of VLMs via Perspective-conditional Prompt-based Contrastive Decoding
Hyejin Chung and Hyunsoo Yoon 21917

HyperTrace: Hypothesis-Based Preference Tracing for Online LLM Personalization
Jianzhi Shen, Keyu Mao, Minghao Shao, Chuanyang Jin, Yusong Wang, Ailiang Lin, Kotaro Funakoshi, Manabu Okumura and Muhammad Shafique 21936

Can LLMs Spot Data Anomalies That Make Table QA Impossible?
Yucheng Wei, Shenzhi Yang, Ning Wang, Jinchang Zhang, Yuhang Yang, Haobo Wang, Liangyu Zha, Wentao Ye, Liyao Li, Cheng Peng and Gang Chen 21952

Uncertainty Is Not a Safety Net for Clinical VQA, but Can It Anticipate Model Failure?
Arnisa Fazla, Alberto Testoni, Ameen Abu-Hanna, Barbara Plank and Iacer Calixto 21988

Agent Planning Benchmark: A Diagnostic Framework for Planning Capabilities in LLM Agents
Haoyu Sun, Wenxuan Wang, Mingyang Song, Jujie He, Weinan Zhang, Yang Liu, Yang Yang and Yu Cheng 22008

TimeClaimBench: A Claim-Level Benchmark for Grounded LLM Explanations of Time Series
Sijia Peng, Fan Zhang, Yixi Zhou, Changlun Li, Yun Xiong, Yangyong Zhu, Xi Chen, Yanwei Yu and Nan Tang 22071

Readable Yet Unpredictable: Rotated-Outcome Prediction in Vision-Language Models
Lexin Wang, Shenghua Liu, Yiwei Wang, Jiafeng Guo and Xueqi Cheng 22095

Modal Sense Disambiguation in Language Models: The Role of Context, Training Domain, and Input Frequency
Ezgi Başar, Yevgen Matusevych, Annemarie van Dooren and Arianna Bisazza 22112

TRACE: A Fine-Grained Benchmark for Temporal Reasoning and Chronological Knowledge Editing over Large Language Models
Xiliang Liu, Zhiwen Xie and Derek F. Wong 22130

SENSESHIFT: Continuous Sentiment-Controlled Text generation via Encoder-based Mask Infilling
Shahed Masoudian, Markus Frohmann, Emmanouil Karystinaios, Navid Rekabsaz and Markus Schedl 22144

FineKB: Asymmetric Multi-Centroid Retrieval for Noisy Case-to-KB Matching in Enterprise Support
Murat Kalender, Aashutosh Nema, Samaksh Gulati, Jeannie M Fitzgerald and Mustafa S. Cetin 22164

Continual Fine-Tuning of Large Language Models via Program Memory
Hung Le and Svetha Venkatesh 22177

Larger Context Window, Fewer Overcorrections: Optimizing Prompts and Batching for Minimal-Edit Grammatical Error Correction
Kateryna Karpo and Artem Chernodub 22194

Towards Efficient Chain-of-Thought Reasoning via Adaptive-Budgeting based Policy Optimization
Wenhao Lin, Qiong Wu, Shuaixiang Yang, Yiyi Zhou, Hui Li, Kehai Chen, Xiaoshuai Sun and Rongrong Ji 22223

Formatting and Surface Feature Confounds in AI Text Detection: A Clever Hans Effect
Koel Dutta Chowdhury, Cristina España-Bonet and Josef Van Genabith 22266

SIN-Bench: Tracing Native Evidence Chains in Long-Context Multimodal Scientific Interleaved Literature
Junjie Wang, Yiming Ren, Yuxin Meng, Yihang Shi, Zhiqiang Lin, Yulin Song, Ruihang Chu, Yiran Xu, Ziming Li, Yunfei Zhao, Zihan Wang, Yu Qiao, Ruiming Tang, Minghao Liu and Yujiu Yang 22278

Deliberation on Demand: Selective Deep Thinking for Language Agents
Junnian Chen, Jianxun Lian, Qinlu Cao, Chunlin Liu, Jiwei Zhang, Mingyang Zhou and Hao Liao 22303

CounterMem: Counterfactual Evaluation of Non-Parametric Memory Use in Retrieval-Augmented Generation
Mikhail Baklanov, Oren Barkan and Noam Koenigstein 22332

ICE: Intervention-Consistent Explanation Evaluation with Statistical Grounding for LLMs
Abhinaba Basu and Pavan Chakraborty 22349

A Shaky Voice Is Not Always a Dodge: Benchmarking Textual and Vocal Evasion Detection in Earnings Calls
Mirae Kim, Seonghun Jeong and Youngjun Kwak 22366

SPSR: Achieving Superior Large Language Model Layer Pruning Performance by Super-fast Recovery
Zhiguo Yang, Changjian Deng, Qinke Chen, Zijing Zhou and Jian Cheng 22378

HyperSU: Corpus-Driven Semantic-Unit Hypergraph for Retrieval-Augmented Generation
Jiate Liu, Liuyi Chen, Zhengyi Yang, Chuan He, Mingchen Ju, Bocheng Han, Ruyi Liu and Xu Zhou 22399

Conditional Perturbation Generators for Probing and Mitigating Recovery Risks in Multimodal Large Language Models Unlearning
Yiqun Kou, Bowen Hu, Siyuan Wang, Erfei Cui, Hongbang Fan, Sizhe He, Yujie Yang and Yadong Zhou 22425

HiGraph: Explicit Discrepancy Reasoning with Hierarchical Graph Propagation for Joint Pronunciation Assessment
Hyunku Kang, Eunseo Choi, Ireh Kim and Chanwoo Kim 22441

ProRAC: A Neuro-symbolic Method for Reasoning about Actions with LLM-based Progression and Search
Haoyong Wu and Yongmei Liu 22456

CapGeo-Bench: Decoupling Visual Perception from Reasoning and Evaluating Geometric Understanding
Yuying Li, Siyi Qian, Hao Liang, Leqi Zheng, Ruichuan An, Linzhuang Sun, Jiajun Zhang and Wentao Zhang 22475

Why This and Not That? A Collaborative Reflection Approach for Understanding Thought Coverage in Decision Making Support Dialog
Morita Tarvirdians, Hayley Hung and Catharine Oertel 22495

Using Grounded Theory for Agent Behavior Analysis at Scale
Zhuoran Lu, Yangyang Yu, Zhuoyan Li, Yibo Meng, Nan Jiang, Chengxi Zang, Jie Gao and Ziang Xiao 22511

D2C-Routing: Dimension-to-Composition Evidence Routing for Mixed-Origin AI-Generated Text Detection
Xin Chen, Fuwei Zhang, Yiqi Tong, Wei Guo, Yutian Xiao and Fuzhen Zhuang 22535

OmniDA: Grounding Agentic Multimodal Data Analysis with Evidence-Aware Process Rewards
Xuhang Zhu, Liyao Li, Ningtao Wang, Yanyu Chen, Hao Chen, Wentao Ye, Ziquan Wang, Haoxing Zhai, Yu Cheng, Gang Chen, Xiu Tang and Haobo Wang 22552

Learning to Think in Blocks: A Prior-Guided Reinforcement Learning Framework for Deep Research
Wangyi Mei, Ganlin Xu, Zhoujia Zhang, Jiaqing Liang, Weijia Lu, Jinhui Huang, Chen Yang, Xiaodong Zhang, Zhifei Yang, Xiaofeng Ma and Deqing Yang 22566

Don't Lose Entities from Retrieval to Generation: Dual Entity Recovery RAG for multi-hop QA
Heechang Lee and Dongyoung Lim 22585

Scalable Supervision for Software Agents via Patch Reasoning
Junjielong Xu, Boyin Tan, Xiaoyuan Liu, Chao Peng, Pengfei Gao and Pinjia He 22603

GD^{2}PO: Group-Dynamic Reward-Decoupled Policy Optimization for Mitigating Multi-Reward RL Conflicts
Haotian Liu, Yihao Liu, Jingwei Ni, Siyuan Huang, Xinpeng Liu, Pengyu Cheng, Jiajun Song, Ruijin Ding, Junfeng Li, Zhechao Yu, Mengyu Zhou, Hongteng Xu, Xiaoxi Jiang and Guanjun Jiang 22624

DailyReport: An Open-ended Benchmark for Evaluating Search Agents on Daily Search Tasks
Jingxuan Han, Wei Liu, Mingyang Zhu, Youpeng Wang, Ziwen Wang, Lin Qiu, Xuezhi Cao, Xunliang Cai, Zheren Fu, Licheng Zhang and Zhendong Mao 22645

JPhonABX: Japanese Phonological Discrimination Dataset for Moraic and Accentual Minimal Pairs
Yurie Koga, Shunsuke Kando and Yusuke Miyao 22676

LLMBridge: An LLM Pipeline for End-to-end Referential Bridging Resolution in English
Lauren Levine and Amir Zeldes 22696

Efficient Diffusion Language Models: A Comprehensive Survey
Haokun Lin, Xinle Jia, Shaozhen Liu, Shujun Xia, Weitao Huang, Haobo Xu, Junyang LI, Yicheng Xiao, Xingrun Xing, Ziyu Guo, Renrui Zhang, Qi Li, Yichen Wu, Renzhen Wang, Xiaojuan Qi, Deyu Meng, Caifeng Shan, Hongsheng Li and Zhenan Sun 22711

Don't Scroll Back: Missing-Evidence Memory for Streaming Dialogue Summarization
Hyangsuk Min and Hwanjun Song 22728

DPA-MoE: A Runtime Precision Control Framework for MoE Inference
Jaeyun Kim, Hoonki Lee, Susie Shin, Hyunmin Kim, Dahyun Lee and Sangyeob Kim 22761

PDOS: Widening the Safety Overton Window Through Persona-Driven Generation
Kritesh Rauniyar, Mark Dras and Usman Naseem 22779

EmoDistill: Offline Emotion Skill Distillation for Language Model Agents in Adversarial Negotiation
Yunbo Long, Haolang Zhao, Lukas Beckenbauer, Liming Xu and Alexandra Brintrup 22804

Confident but Wrong: A Constrained Decoding Diagnostic for Low-Resource Automatic Post-Editing
Isuru Wijesiri, Nisansa de Silva, Kavindu Warnakulasuriya, Aloka Fernando and Surangika Ranathunga 22838

Do Vision-Language Models Understand Visual Persuasiveness? A Diagnosis via Visual Persuasive Factors
Gyuwon Park and Hyounghun Kim 22867

VEN-VL: A Visual Ensemble MoE Framework for Effective and Efficient Multi-Modal Understanding
Yinghao Wu, Zhuoyan Luo, Yiyao Yu, Zhaojian Yu, Yujiu Yang and Xiao-Ping Zhang 22906

A Unified Per-Token Gating Family for On-Policy Distillation: FKL/RKL Mixing with Multi-Channel and Bias Coefficients
Suwan Wu, Yumeng Lin, Pengcheng Yuan and Xiaolong Jiang 22921

Knowing but Not Doing: Discourse-level Constraints Adherence of Large Language Models
Zhonghua Hong, Sihang Jiang, Keyi Wang, Xiaojun Meng, Jiansheng Wei, Jiaqing Liang and Yanghua Xiao 22934

ASP: Adaptive Semantic Preservation for Context Compression
Shaoshen Chen, Peiyang Li, Jiaming Mu, Guixiong Chen, Fengting Li and Chao Feng 22950

SCOPE-RL: Optimizing Reasoning Paths Before and After Success
Xiaojian Liu, Han Xu, Jianqiang Xia, Zhixuan Li, Ke Xu, Yiwei Dai, Xinran Chen, Changwo Wu and Yuchen Li 22968

Generating Complete Court Judgments with Large Language Models: A Study on CMML
Zhongyi Guan, Zhitian Hou, Nanli Zeng and Kun Zeng 22990

Dive: Accurate and Efficient Video Large Language Models for Lifespan Question via Head-wise Visual KV Cache Compression
Sanghwa Hong, Jongjin Kim and U Kang 23002

Reinforcing Traceable Medical Reasoning with Critical Evidence Graph
Linjie Mu, Yannian Gu, Zhongzhen Huang, Yakun Zhu, Shaoting Zhang and Xiaofan Zhang 23019

Bridging the English-Arabic Medical Knowledge Gap: Targeted Low-Rank Adaptation via Causal Layer Selection
Chaimae Abouzahir, Musa Khan, Hala Ali-Hassan, Congbo Ma, Khaled Saleh, Yousra Sadqi, Jihad Mallat, Walid Al-Eisawi, Nizar Habash and Farah E. Shamout 23043

TUX: Measuring Human–AI Tacit Understanding
Yueshen Li, Hanyi Min, Vedant Das Swain and Koustuv Saha 23083

PSR: Efficient Knowledge Graph Question Answering via Plan-Guided Schema-Level Reasoning
Mingyuan Li, Yucheng Li, Yuheng Bao, Yawei Sun and Gong Cheng 23104

SCENE: A Scenario Completion Benchmark for Measuring Social Bias in Large Language Models
Migyeong Kang, Aecheon Jung, Eunchae Jang, Minseok Kim, S. Shyam Sundar, Sungeun Hong, YunSeok Choi and Jinyoung Han 23126

Performance of Grapheme-Based Tokenizers on Word-level and Natural Language Understanding tasks for Indic Languages
Shagun Dwivedi, Mohammad Saquib and Kaushik Gopalan 23149

A MultiModal Autoregressive Transformer Framework for Social Popularity Prediction
Mingzhu Wang, Dong Zhou and Zeliang Zhou 23162

From Separation to Entanglement: Dissecting Feed-Forward Network Functional Organization in Vision–Language Models for Ancient Chinese Documents
Ruilin Liu, Dongmei Zhu, Yuting Wei and Jin Song 23176

When and What to Teach: Budget-Aware Online Adaptation for Web Agents
Jianwei Zhang, Sihan Cao, Pengcheng Zheng, Ya Wen, Pei Ke, Kuien Liu, Shen Gao, Wei Dong, Yang Yang and Chaoning Zhang 23190

Mining or Synthesis? Rethinking Exploration Efficiency in Iterative Alignment of Mathematical Reasoning
Jun Rao, Zixiong Yu, Xuebo Liu, Guhan Chen, Jing Li, Hejin Wang, Jiansheng Wei, Xiaojun Meng and Min Zhang 23206

Scientific Knowledge-driven Decoding Constraints Improving the Reliability of LLMs
Maotian Ma, Zheni Zeng, Zhenghao Liu and Yukun Yan 23230

HSPMark: Desynchronization-Resilient Multi-Bit LLM Watermarking via Semantic Anchoring
Jing Wang, Pengyu Xiong, Kejun Qian, Yi Yang and Jing Fan 23245

GUI Agents for Continual Game Generation
Yixu Huang, Bo Li, Li Na, Zhe Wang, Kaijie Chen, Haonan Ge, Qingyi Si, Yuanzhe Shen, Ruihan Yang, Guangjing Wang and Hongcheng Guo 23260

MedEQ: A Clinical Emotional Intelligence Benchmark and Rubric-Guided Reinforcement Learning for Patient-Centric Medical LLMs
Yiming Zhang, Xiaoyan Yang, Yanchao Xu, Jiayi Deng, Yingfan Ma, Lirong Gao, Yanmei Gu, Zhengkaiya, Bowen Song, Jinjie Gu and Junbo Zhao 23279

AgenTRIM: Tool Risk Mitigation for Agentic AI
Roy Betser, Amit Giloni, Shamik Bose, Sindhu Padakandla, Chiara Picardi, Lidor Erez and Roman Vainshtein 23301

Why Prompt-Injection Detectors Fail on XPIA: A Latent-Space Explanation
Yash Narendra and Niladri Chatterjee 23326

DIAG: Diagnostic Iterative Alignment and Generation for Data-Efficient Mathematical Preference Distillation
Guhan Chen, Songtao Tian, Bohan Li, Hejin Wang, Ye-Xin Xie and Zixiong Yu 23345

Multilingual Medical Reasoning for Question Answering with Large Language Models
Pietro Ferrazzi, Aitor Soroa and Rodrigo Agerri 23365

VeriTrace: Evolving Mental Models for Deep Research Agents
Haolang Zhao, Yunbo Long, Lukas Beckenbauer and Alexandra Brintrup 23388

LLM Knows, Speaks Not: Extracting Decoder-State Representations for Column Type Annotation
Lingxi Cui, Jianfeng Zhang, Huan Li, Ke Chen and Lidan Shou 23427

CompassJudger-2: A Holistic Approach Towards Generalist Judge Model
Taolin Zhang, Zihan Ma, Maosongcao, Shudong Liu, Junnan Liu, Songyang Zhang and Kai Chen 23442

BioELX: Context-Aware Cross-lingual Biomedical Entity Linking without Task-Specific Supervision
Yi Wang, Corina Dima, Liangyu Zhong and Steffen Staab 23462

Med-Chain: An Adaptive Clinical-Chain Retrieval Framework for Evidence-Grounded Medical Question Answering
Fucheng Niu, Songtao Jiang, Qinxin Wu, Sylvia Chung Yan Shan, Tianxiang Hu, Jian Wu and Zuozhu Liu 23489

HW-GNN: Homophily-Aware Gaussian-Window Constrained Graph Spectral Network for Social Network Bot Detection
Zida Liu, Bao Xiaoyuan, Ji Zhang, Zhao Li and Jun Gao 23509

From Proprietary to Public: Dataset Declassification for Question Answering
Stefano Campese, Alessandro Moschitti and Ivano Lauriola 23526

TimeSage-EV: A Live Benchmark for Agentic Time Series Analysis in Evolving Environments
Qingren Yao, Yaxuan Kong, Yuqi Nie, Yichen Li, Stefan Zohren, Anna Vettoruzzo, Qingsong Wen, Ming Jin and Joaquin Vanschoren 23542

Robust Code RL via Faulty-Code-Driven Test case Synthesis and Dense Reward Shaping
Yiwen Zhang, Xiaodong Yan, Zhenyu Huang, Deng Zhao, Liang Jiang, Qing Cui, Zujie Wen, Zhiqiang Zhang and Jun Zhou 23568

Is Multilingual LLM Watermarking Truly Multilingual? Scaling Robustness to 100+ Languages via Back-Translation
Asim Mohamed and Martin Gubri 23589

RIFT-Bench: Dynamic Red-teaming For Agentic AI Systems
Yarin Yerushalmi Levi, Roy Betser, Amit Giloni, Lidor Erez, Itay Gershon, Oren Rachmil, Sindhu Padakandla and Roman Vainshtein 23617

λ-Diffusion Language Models: Controllable Generalization of Continuous and Discrete Diffusion Dynamics
DongNyeong Heo, Hyunmin Song and Taehwan Kim 23666

AdaMemPIP: Adaptive Memory Pipeline Orchestration for LLM Agents
Huicheng Huang, Xiaohui Zhang, Yuanning Cui, Zequn Sun and Wei Hu 23684

Plan, Verify and Fill: A Structured Parallel Decoding Approach for Diffusion Language Models
Miao Li, Hanyang Jiang, Sikai Cheng, Hengyu Fu, Yuhang Cai, Baihe Huang, Zhongzhu Chen, Tinghan Ye, Xuanzhou Chen and Pascal Van Hentenryck 23694

Improving Speech Generation via Multi-Neighbor Token Prediction for Neural Codec Language Models
Wenrui Liu, Qian Chen, Wen Wang, Yuxuan Wang, Haoxiang Wu, Guanrou Yang, Rui-Chen Zheng, Weiqin Li, Guo Tao, Xiangang Li and Zemin Liu 23718

Modular Expert Merging for Biomedical Retrieval
Sameh Khattab, Jean-Philippe Corbeil, Osman Alperen Çinar-Koraş, Amin Dada, Julian Friedrich, Jiawei He, Douglas Teodoro and Jens Kleesiek 23732

PerturbPrune: Guiding Sparsity Allocation by Probing Layer Stability Along Principal Directions
Ningxin Wu, Zhengxu Hou, Ziran Zhao, Tao Jing, Pengyuan Liu, Chen Kang, Yangshijie Zhang, Luxi Xing, Bin-Bin Yang, Feng Xiao, Lu Liu and Dong Yu 23753

Compile, Don't Memorize: A Context Compilation Architecture (CCA) for In-Context Learning
Jinhu Qi, Minda Hu, Wentao Zhang, Weiqiang Jin, Yanyu Chen, Junli Wang and Irwin King 23781

Adversarial Robustness of Activation Steering in Large Language Models
Kien Le and Thai Le 23804

AeroVerifier: A Typed Spatial Verifier for Frozen Vision-Language Models in Aerial Visual Question Answering
Jinfeng Chen, Mengfan Yang, Jianyu Zhu, Zhicheng Zhou and Jing Liang 23826

Exploring Language-Agnosticity in Function Vectors: A Case Study in Machine Translation
Nurkhan Laiyk, Gerard I. Gállego, Javier Ferrando and Fajri Koto 23840

SWITCH: Benchmarking Modeling and Handling of Tangible Interfaces in Long-horizon Embodied Scenarios
Juntao Cheng, Wanyue Zhang, Zhiwei Yu, Shuo Ren, Zheqi He, Shaoxuan Xie, Guocai Yao, Jieru Lin, Börje F. Karlsson and Jiajun Zhang 23852

Every Step Counts: Step-Level Credit Assignment for Tool-Integrated Text-to-SQL
Yaxun Dai, Baolin Sun, Junying Wang, Pengfei Wang, Yingqi Gao, Xuemei Dong, Mengdie Chu, Xiang Qi and Pingfu Chao 23876

StanceAgent-R1: Agentic Reinforcement Learning for Zero-Shot Conversational Stance Detection
Xing Li, Jinzhong Ning and Yijia Zhang 23895

Reflection Is Not Monolithic: Experience-Driven Control of Recheck in LLM Reasoning
Quanyu Long, Jiang Kai Jie, Jianda Chen, Qisheng Hu, Xu Guo, Leilei Gan and Wenya Wang 23920

SYNAPSE: Neuro-Symbolic Visual Thought-to-Text Decoding via Topological Semantic Denoising
Akshaj Murhekar and Abhijit Mishra 23941

RecToolBench: Benchmarking Recommendation-Specific Tool Orchestration under Fuzzy User Intent
Xiao Chen, Yicheng Zhao, Yingying Wu, Zhendong Chu, Changyi Ma, Qingsong Wen and Xuan Song 23958

Are Large Language Models Suitable for Graph Computation? Progress and Prospects
Yuting Zhang, Yi Han, Kai Wang, Wei Ni, Angela Bonifati and Wenjie Zhang 23984

EvalMem: An Operation-Level Diagnostic Framework for Long-Term Memory Systems
Zeyu Liu, Jian Zhong, Rongduo Han, Ziyang Wu, Shunye Tang, Chenghao He, Yaxuan Yang, Yihang Qiu, Ailing Wang, Xiao Liang, Guohuan Xie, Xiaokang Xue, Gongchen Li, Haining Zhang and Wei Wang 24002

Probe to Act: Elevating Browser-Use Agent via Active Visual Probing
Keliang Li, Heng Wang, Chen Hu, Daxin Jiang, Hong Chang and Shiguang Shan 24035

EgoActor: Grounding Task Planning into Spatial-aware Egocentric Actions for Humanoid Robots via Visual-Language Models
Yu Bai, MingMing Yu, Chaojie Li, Ziyi Bai, Xinlong Wang and Börje F. Karlsson 24059

Difficulty-Adaptive Tree-Structured Policy Optimization for Expanding Reasoning Coverage in RLVR
Youngjun Yu, Sanghwan Jang and Hwanjo Yu 24084

PACE: A Unified Condense-and-Extract Paradigm for Fast VLM Inference
Junjie Liu, Shengyuan Ye and Xu Chen 24111

Where Identity Lives: Localized, Retain-Free Identity Unlearning in Multimodal Large Language Models
Kangwook Ko, Jaehyuk Jang, Wonjun Lee, Hee-Seon Kim and Changick Kim 24133

Beyond Trajectory Rewards: Step-level Credit Assignment for Agentic Search via Graph Modeling
Yuchen Liu, Feng Yingjie, Lixiong Qin, Jiasi Chen, Jianing Yu, Sheng Gao, Sheng Yang and Weiran Xu 24150

When Not to Imitate: Boundary-Aware Skill Memory for Reliable Tool-Use LLM Agents
Zihan Lin, Zhenyu Chen, Jiawen Wei, Xiaohan Wang, Jie Cao, Jiajun Chai, Wei Lin, Guojun Yin and Ran He 24166

Remembering Facts, Updating Minds: Personal-Factual Adaptive Memory for Multi-Turn Emotional Support
Hao Wang, Qiuyun Zhang, Peilin LI and Ke Ma 24183

Evaluate-as-Action: Self-Evaluated Process Rewards for Retrieval-Augmented Agents
Jiangming Shu, Yuxiang Zhang, Ye Ma, Xueyuan Lin and Jitao Sang 24206

SFAD: Speculative Factuality-Aware Decoding
Guanqiao Chen, Di Wang and Lijie Hu 24221

JaggedBench: Measuring Reasoning Invariance via Automated Weakness Exploration
Donghao Luo, Chenchen Zhang, Siheng Li, Zenan Xu, Qi Yi, Boyuqiu, Weikang Wang, Guanhua Huang, Tao Yang, Bo Zhou and Alex Chan 24240

BDI-batch: Leveraging Standardized Clinical Questionnaires for Contrastive Learning in Psychological Marker Retrieval
Marcos Fernández-Pichel and David E. Losada 24280

SemFed: Semantic Federated Reinforcement Learning for Heterogeneous LLM Agents
Yuxuan Luan, Zilin Zhao, Yihan Zhang, Wei Wang, Chenyu Sun, Lixiang Li, Haipeng Peng, Chaoning Zhang and Lianku Xie 24294

STAGE: Diagnosing Semantic Transfer at Grounded Execution in Embodied Agents
Baosheng Jin, Yushen Liang and Hua Shen 24309

Asking the Right Questions: Ontology-Grounded Interpretable Embeddings for Biomedical Text
Yixuan Tang, Zhenghong Lin, Yandong Sun, Wynne Hsu, Mong-Li Lee and Anthony Kum Hoe Tung 24332

Diagnosing Decision-Boundary Induction in Few-Shot In-Context Learning
Mingyang Li, Xinlang Yue, Muyang Li, Yongqi Liu and Kaiqiao Zhan 24345

InSight-doc: Agentic Visual Perception for Long-Document Understanding
Kaican Li, Weiyan Xie, Lewei Yao, Jiannan Wu, Lanqing Hong, Yongxiang Huang and Nevin L. Zhang 24359

Tracing Query Expansion Effects through Sparse Autoencoder Features
Fangan Dong, Weiran Shi, Zhiwei Xu, Xuri Ge, Ben He, Xin Xin, Zhumin Chen and Ying Zhou 24390

A Controlled Study of Auxiliary Supervision for Low-Resource Emotion Classification
Martina Szabóová, Daniel Zemančík and Ján Magyar 24407

DeepDiver-DR: Empowering LLMs for Deep Research via Specialized Multi-Agent Collaboration
Zengyan Liu, Han Wu, Haonan Xu, Wenxuan Shi, Jingyi Wang, Haochen Tan, Jiahui Gao, Haoli Bai, Xiaoguang Li and Linqi Song 24421

On Large Language Models for Effective Red Teaming
Ming Dong, Hongjing Ren, Bolong Zheng, Chengji Wang, Jingwen Cai, Guanyi Chen and Tingting He 24436

Beyond Visual Neglect: Context-Preference Activation Steering for Mitigating MLLM Hallucinations
Jingwen Wu, Xijun Zhang and Ge Song 24454

Evaluating Demographics Effect in Policy-Governed Tool-Using Agents Across Real-World Service Domains
Ilias Driouich, Hongliu Cao and Eoin Thomas 24471

VisMind: Stateful Synthetic Dialogue for Visual Mid-Training
Xinlong Fu, Zihao Zhou and Qiufeng Wang 24493

EDRAC: Benchmarking Arabic Dialect Reading Comprehension
Noor Abo Mokh, Kirill Chirkunov, Teresa Lynn, Nizar Habash, Reham Marzouk, Malik H. Altakrori, Younes Samih, Muhammed AbuOdeh, Nour Rabih, Rahaf Ashahrani, Hamad Alshehhi, Hamdan Al-Ali, Muhra AlMahri, Besher Hassan, Mohamed Anwar, Abed Alhakim Freihat, Preslav Nakov and Alham Fikri Aji 24511

RoleArena: Autonomous Plot Progression for Long-Horizon Multi-Agent Role-Playing
Qi Zhao, Jian Li, Yizhang Jin, Dongqi Liu, Hang Ding, Qi Song and Xiang-Yang Li 24553

DARTS: Diversity-Aware Reasoning Tree Search for Test-Time Scaling
Kaifeng Liu, Shijie Wang, Yuanxing Liu, Yang Liu, Xinyi Wang, Haozheng Hou, Biqing Qi, Weinan Zhang and Ting Liu 24577

Procedural Blind Spots in Agent Evaluation: A Survey, Coverage Analysis, and New Missions for NLP
Hongliu Cao, Ilias Driouich and Eoin Thomas 24601

Arabic Sentence Segmentation Across Genres and Punctuation Conditions
Mohammed Elkholy, Khalid N. Elmadani, Nizar Habash and Bashar Alhafni 24634

AutoBackdoor: Automating Backdoor Attacks via LLM Agents
Yige Li, Zhe Li, Wei Zhao, Nay Myat Min, Hanxun Huang, Xingjun Ma and Jun Sun 24653

Prompted Values and Deployed Policies: A Truth–Propriety Diagnostic in Language Models
Ziyuan Feng, Ajian Liu, Hui Ma and Yanyan Liang 24673

When Reward-Free Signals Are Not Directions: A Gradient-Level Diagnostic for Test-Time Reasoning Optimization
Shijie Zheng, Yinxing Dai and Zhiqiang Pu 24687

Scope3Trace: Evidence-Based Identification and Extraction of Scope 3 GHG Emissions from Sustainability Reports
Siyuan Zheng, Yifan Duan, Chao Xue and Flora D. Salim 24704

STRETCH the Boundaries: A Unified Self-Taught Framework for Progressive LLM Evolution
Yajie Yu, Mark Lee and Yue Feng 24732

A Unified Framework for Low-Resource Sawndip–Chinese Bidirectional Translation and Bilingual Dialogue Modeling
Zhengyuan Wang, Yixin Wu, Jingwen Xu, Zisu Huang, Changze Lv, Xiaohua Wang, Yiyang Lu, Muzhao Tian, Xiaosheng Pan, Menghan Zhang, Xuanjing Huang and Xiaoqing Zheng 24748

BenchCompass: From Scores to Signals for Training and Harness Decisions in Payment-Domain LLMs
Sijie Dong, Wei Ren, Xuanwei Hu, Jiawei Luo, Zifan Wang, Xiaoyun Feng, Hui Cai, Lyuxin Xue, Peng Lu, Wei Wu and Xin Zhang 24762

F^2DR: A Fine-Grained Full-Pipeline Reward Framework for DeepSearch Workflows
Bojian Xiong, Wentao Ding, Yujing Lu, Shaowei Zhang, Ling Shi, Jing Liao, Yan Wang, Yueyang Zhang, Long Xia, Zhiyuan Sun, Daiting Shi, Jingzhou HE, Yuqi Ren and Deyi Xiong 24782

Teacher-Guided Policy Optimization for On-Policy Reasoning Distillation under Large Policy Divergence
Xinyu Liu, Kechen Jiao, Chunyang Xiao, Runsong Zhao, Junhao Ruan, Bei Li, Jiahao Liu, Qifan Wang, Xin Chen, Jingang Wang, Chenglong Wang, Tong Xiao and JingBo Zhu 24801

Script Choice in LLMs: Evidence for Late-Layer Commitment
David Kletz, Sandra Mitrovic, Itay Sabato, Ljiljana Dolamic and Fabio Rinaldi 24817

IPA-Verified Singable Lyrics Translation with LLM Agents
Guan-Ming Chiu, Cheng Chiao-Chih and Kuan-Wei Lee 24826

Argument Quality Assessment with Large Language Models: A Pairwise Bradley-Terry Approach
Nicolás Benjamín Ocampo, Agnes Paullate Nyiranziza and Davide Ceolin 24842

Diagnosing and Selecting Synthetic Motion Tasks for Video RLVR
Yishu Lei, Aofan Liu, Danxiang Zhu, Canming Ye, Dan Zhang, Songhe Deng, Xianlong Luo and Shikun Feng 24858

Beyond Alignment: Value Diversity as a Collective Property in Multicultural Agent Systems
Shaoyang Xu, Jingshen Zhang, Long P. Hoang, Jinyuan Li and Wenxuan Zhang 24872

Syn-Omni: Structured Specialization and Progressive Collaboration for Omnimodal Embeddings
Youngtaek Oh, Qiyu Wu, Hiromi Wakaki, Junmo Kim and Yuki Mitsufuji 24889

ReFrame: Evidence-Guided Test-Time Safety Alignment in Multimodal Large Language Models
Wenzheng Jiang, Xuankun Rong, Yuanzhao Zhai, Feng Dawei and Huaimin Wang 24916

SMAH-CTI: A Unified Sentence-Level Benchmark for Multi-Task Cyber Threat Intelligence Extraction
Niveen Omar Jaffal, Radi Jarrar, Mohammed Alkhanafseh and David Mohaisen 24937

Evaluating Large Language Models on Urdu Idioms
Muhammad Farmal Khan and Mousumi Akter 24956

Can LLM Annotators Match Human Consistency? Benchmarking Human Agreement and Optimizing Multi-Agent with Contrastive Knowledge
Xiaopeng Ke, Zhixin Wang, Xuebo Liu, Xinyu Ma, Xing Zheng, Bowen Song, Weiqiang Wang and Min Zhang 24970

GRAM: Empowering Agent with Actively Managed Graph-Structured Memory via Reinforcement Learning
Xiangqi Chen, Zihao Yi, Haohao Luo and Ying Shen 25004

When Audit Quality Fails to Predict Downstream Utility: Synthetic-Data Selection for Low-Resource Classification
Xuan-Bach Le, Phat T. Tran-Truong and Son Ha Xuan 25027

Attention-Aligned Reasoning for Large Language Models
Hongxiang Zhang, Yuan Tian and Tianyi Zhang 25046

Attribute-Based Diagnosis of LLM Alignment with Hate Speech Annotations
Mohamed Amine Jradi, Faeze Ghorbanpour and Alexander Fraser 25065

Decoding Self-Knowledge from Latent Reasoning Models for Selective Prediction and Adaptive Routing
Zhiqi Li, Yichi Zhang, Cunjie He, Ziheng Hua, Maolin Wang and Zenglin Xu 25076

Beyond Paraphrases: Generalization Across Task Formats in Parametric Knowledge Editing
Cheongwoong Kang, Jongeun Baek, Yeonjea Kim and Jaesik Choi 25089

RankSteer: Can Pointwise LLM Rankers Be Calibrated at the Representation Level?
Yumeng Wang, Catherine Chen and Suzan Verberne 25111

PRISM: Calibrated Prior-Risk Integration for Safe and Useful Test-Time Alignment
Yiqi Deng, Ruixi Liu, Jiahao Hu and Guoai Xu 25128

SAGE: An Extensible Benchmark for Service Agent Graph-guided Evaluation
Ling Shi, Yuqin Dai, Ziyin Wang, Wei Zhang, Ning Gao, Chaozheng Wang, Xinwei Wu, Yujie Wang, Wei He, Jinpeng Wang, Linlong Xu and Deyi Xiong 25145

Context-anchored Asymmetric Conformal Prediction for Uncertainty-aware Reasoning over Knowledge Graphs
Boyang Fan, Ziyue Qiao, Wei Ju and Xiao Luo 25165

AdaptiveGuard: Treating Constraints as Falsifiable Hypotheses for LLM Agents in Dynamic Interactive Environments
Pengxiang Liu, Tao Ren, Wei Xiong, Lijie Li, Junjie Wang and Jun Hu 25186

Preserving Memory Function in Hybrid Language Models: A Survey of State, Evidence, Access, and Granularity
Zhuangfei Hu, Runqing Xu, Yuchen Li and Yongqi Zhang 25206

ExecRubrics: Executable Tool-Augmented Rubrics for Verifiable, Efficient and Explainable Evaluation
Kaustubh Dhole, Charles L. A. Clarke and Eugene Agichtein 25232

Endogenous Observations in Agent Self-Evolution: A Survey
Zhixin Zhang, Xinke Jiang, Jiaran Gao, Zhibang Yang, Yuheng Huang, Junfeng Zhao, Xu Chu and Yasha Wang 25257

Rethinking Fairness in LLM-Based Recommender Systems: A Survey
Song-Duo Ma, Chu Yun Chen, Bang-An Li, Pin-Yu Chen, Shau-Yung Hsu and Yun-Nung Chen 25277

Enhancing Pun Meme Understanding with Visual Semantic Cues
Meng Li, Zhongqing Wang and Guodong Zhou 25297

SubtleMemory: A Benchmark for Fine-Grained Relational Memory Discrimination in Long-Horizon AI Agents
Wenxuan Wang, Haoyu Sun, Fukuan Hou, Mingyang Song, Weinan Zhang, Yu Cheng and Yang Yang 25309

Beyond Endpoint Gains: A Weight-Delta Audit of Medical Specialization
Praphul Singh, Shanu Kumar and Akshat Agarwal 25360

RelayS2S: Dual-Path Speculative Generation for Real-Time Dialogue
Long Mai and Junli Liang 25379

EDIT: Evidence-Diagnosed Intervention Training for Rule-Faithful LLM Grading
Zhihao Wu, Linhai Zhang, Taiyi Wang, Runcong Zhao, Peter Andrews, Cesare Aloisi and Yulan He 25397

Action-Level Credit Assignment for Retrieval-Augmented Reasoning
Laura Perez-Beltrachini and Mirella Lapata 25414

LLaVA-RadZ: Can Multimodal Large Language Models Effectively Tackle Zero-shot Radiology Recognition?
Bangyan Li, Wenxuan Huang, Zhenkun Gao, Yeqiang Wang and Yuxiang Shen 25435

SEA-LION-Embedding: Open and Reproducible Text Embeddings for Southeast Asia
Peerat Limkonchotiwat, Raymond Ng, Sarana Nutanong and Jian Gang Ngui 25455

Latent-Space Intervention for Cross-Lingual Factual Consistency: Consistency Improvements without Accuracy Drops
Faeze Ghorbanpour, Constanza Fierro, Alexander Fraser and Anders Søgaard 25464

BIRDS: Characterizing and Understanding Biodiversity Impact of Large Language Model Serving
Tianyao Shi and Yi Ding 25476

Sycophancy Suppression Can Impair Rational Updating: Anti-Sycophancy Should Preserve the Ability to Update
Huanhuan Ma, Henry Peng Zou, Chengze Li, Enze Ma, Yunyue Su and Philip S. Yu 25499

OpenIDC: Reconstructing the Evidence Chain for Lightweight Image Difference Captioning
Jiabao Du, Saihui Hou, Yuan Liu, Jiawei Wang and Yongzhen Huang 25516

The Evolutionary Trajectory of Medical Multi-Agent Systems: A Survey
Kai Chen, Chenzhang Li, Tianyi Yu, Changzhuo Zuo, Xinyun Zhou, Sijia Jing, Yanqing Wu, Wenfei Gao, Xinfeng Li, Tianpei Yang, Bo An and Yang Gao 25530

MEMORA: Embodied Action Memory from Egocentric Videos for Reasoning and Planning
Zihao Yu, Xiu Yuan and Chongjie Zhang 25551

GRACE: Gradient-guided Coreset Selection for LLM Unlearning
Praveen Bushipaka, Andrea D'Angelo, Lucia Passaro and Tommaso Cucinotta 25601

SocialMaze: A Benchmark for Evaluating and Enhancing Social Reasoning in Large Language Models in Complex Social Environments
Zixiang Xu, Yanbo Wang, Yue Huang, Haomin Zhuang, Yujun Zhou, Jiayi Ye, Sixian Li, Zirui Song, Lang Gao, Chenxi Wang, Zhaorun Chen, Wang Pan, Yue Zhao, Jieyu Zhao, Xiangliang Zhang and Xiuying Chen 25621

SAGE: Answer-Conditioned Uncertainty Targets for Verbal Confidence Alignment
Kaiwen Shi, Zheyuan Zhang and Yanfang Ye 25717

Don't Box Me In: Dynamic Cultural Adaptation and Cognitive Tracking for Social Understanding
Chongyuan Dai, Yaling Shen, Shengeng Tang, Hui Ma and Jinpeng Hu 25738

Active Learners as Efficient PRP Rerankers
Jeremías Figueiredo Paschmann, Juan Kaplan, Francisco Nattero, Santiago Mauricio Barron Bucolo, Juan Wisznia and Luciano Del Corro 25760

MSQA: A Natively Sourced Multilingual and Multicultural SimpleQA Benchmark
Yukai Huang, Xianru Chen, Mingxiang Chen, Xinping Lei, Fangbing Deng, Jin Chen, Jiaheng Liu, Ge Zhang and Wenhao Huang 25777

UniData: Universal Multimodal Instruction Generation Pipeline
Jiaqi Tang, Yi-Feng Wu, Yuting Zhang, Hao LU, Bowen Fu, Qing-Guo Chen, Xiaogang Xu, Yuwei Hu, Shiyin Lu, Wei Wei, Lei Zhang, Zhao Xu, Weihua Luo, Qifeng Chen and Ying-Cong Chen 25793

When Evidence Shapes Collaboration: Knowledge-Conditioned Topology Generation for Multi-Agent Systems
Yangxiao Jiang, Jiarun Fan, Mingcong Xu, Yanxi Guo, Jiwen Feng, Shanqing Xu, Mengchen Qian, Wei Chen and Xiaojin Zhang 25814

Meta-Learning Where to Allocate Experts: Task-Conditioned Layer-Wise Compression for MoEs
Wang Rongfeng, Shichao Weng, Zhiqiang Wang, Xinyu Liu, Yang Yi, Peilong Zhou and Tang Hongwei 25830

From Complexity to Clarity: A Systematic Study of Geometric Dynamics in Hidden Representations of LLMs
Xiang Li, Luzhe Sun, Matthew Walter and Jiawei Zhou 25843

LongTraceRL: Learning Long-Context Reasoning from Search Agent Trajectories with Rubric Rewards
Nianyi Lin, Jiajie Zhang, Lei Hou and Juanzi Li 25862

LoopGuard: Breaking Self-Reinforcing Attention Loops via Dynamic KV Cache Intervention
Dongjie Xu, Hao Wu, Weijie Shi, Yue Cui, Yuanjun Liu, Jiawei Li, Haolun MA, An Liu, Jia Zhu and Jiajie Xu 25883

Loci Similes: A Benchmark for Extracting Intertextualities in Latin Literature
Julian Schelb, Michael Wittweiler, Marie Revellio, Barbara Feichtinger and Andreas Spitz 25899

CompCQR: Compositional Query Generation for Training-Free Conversational Search
Yunah Jang, Kang-il Lee, Joongbo Shin and Kyomin Jung 25925

GRASP: GRanularity-Aware Search Policy for Agentic RAG
Varun Gandhi, Jaewook Lee, Shantanu Todmal, Franck Dernoncourt, Ryan A. Rossi, Andrew Lan and Zichao Wang 25944

Unsupervised Domain Specialization for Multimodal Embeddings
Ruifeng Hu, Jianxun Lian, Mingqi Wu, Mingyang Zhou, Hao Liao and Naipeng Chao 25976

Conservative Inference-Time Tuning with an Improvement Guarantee
Daiki Chijiwa and Taku Hasegawa 25995

DataCross: A Unified Benchmark and Agent Framework for Cross-Modal Heterogeneous Data Analysis
Ruyi Qi, Zhou Liu, Yang Gu, Zeli Su, Pan Zewei, Xing Chen, Yuanfeng Song, Wangxiaoyue and Wentao Zhang 26006

CoDR: Training-Free Confidence-Drift Remasking for Diffusion Language Models
Yue Wu, Qinghe Zheng, Yu Zhang and Jian Huang 26028

Where LLM Annotators Fail: Label-Free Learning on Graphs with LLMs
Safal Thapaliya, Jiatan Huang and Chuxu Zhang 26045

M-GOAL: Toward Accurate and Language-Consistent Reasoning in LLMs via English-Only Post-Training
Dingyuan Dai, Yuxi Chen and Jiayuan Wang 26066

Chain-of-Trigger: An Agentic Backdoor that Paradoxically Enhances Agentic Robustness
Jiyang Qiu, Xinbei Ma, Yunqing Xu, Zhuosheng Zhang and Hai Zhao 26085

FactNLI: Dynamic and Automated Fact-Based Augmentation of NLI Benchmarks
Yihan Lu, Junzhe Zhang, Huixuan Zhang and Xiaojun Wan 26107

GLIDE: Generalized Layer-wise Intrinsic Distributional Evaluation for Heterogeneous LLM Agents
Wei Zhu, Yiming Wang, Rui Wang, Lixing Yu, Kun Yue and Zhiwen Tang 26128

MemoNoveltyAgent: A Historical-Memory-Augmented Multi-Agent System for Paper Novelty Assessment
JiaJun Hou, Hexuan Deng, Wenxiang Jiao, Xuebo Liu, Xiaopeng Ke, Derek F. Wong and Min Zhang 26159

Agora: Enhancing LLM Agent Reasoning Via Auction-Based Task Allocation
Kaiji Zhou, Ales Leonardis and Yue Feng 26189

Quantifying Self-Preservation Bias in Large Language Models
Matteo Migliarini, Joaquin Pereira Pizzini, Luca Moresca, Valerio Santini, Indro Spinelli and Fabio Galasso 26202

Speculative Probing: LLM Monitoring at Speculative-Decoding Cost
Collin Zhang, Tingwei Zhang and Vitaly Shmatikov 26227

VC-IF: Introducing Visual-Constraint Instruction Following for Multi-Turn, Multi-Constraint Vision-Language Models
Jixiang Hong, Pi Bu, Yingyao Wang, Bo Peng, Peijie Wang, Chen Wang, Tengtao Song, Yang Yao, Jun Song, Rui Yan, Cheng Yu and Bo Zheng 26242

KLOD: Locality-Preserving Knowledge Editing via Non-Target Distribution Preservation
Hojun Jeong, Gyunyeop Kim and Sangwoo Kang 26264

Observation-Aware Stabilization of Reinforcement Learning for Retrieval-Augmented Generation
Naixin Zhang and Tong Lin 26287

FormalImG: Evaluating Structural Compositional Generalization for T2I Models
Hong-Jie You, Jie-Jing Shao, Xiao-Wen Yang, Zhi-Fan Wu, Lin-Han Jia, Lan-Zhe Guo and Yu-Feng Li 26308

ACRE: An Acoustic-Cue and Contextual-Knowledge Reasoning Enhancement Framework for Speech Relation Extraction
Wenxuan Mu, Jinzhong Ning, Di Zhao and Yijia Zhang 26343

Beyond Static Summarization: Proactive Memory Extraction for LLM Agents
Chengyuan Yang, Zequn Sun, Wei Wei and Wei Hu 26353

Abstraction Agent: LLM-Guided Abstraction for Imperfect-Information Games
Boning Li and Longbo Huang 26368

PUMA: Post-Hoc Sparsification of Universal Multimodal Embeddings for Efficient Retrieval
Matteo Attimonelli, Alessandro De Bellis, Franco Maria Nardini, Claudio Pomo, Cosimo Rulli, Rossano Venturini and Tommaso Di Noia 26388

Polite but Misaligned: Evaluating LLM Politeness Judgments Against Human Pragmatic Norms
Rong Wang, Kun Sun and Yadong Guo 26402

From Final Artifacts to Trajectories: Retrospective Process Supervision for Evidence-Grounded Long-Form Generation
Junjie Huang, Jiarui Qin, di Yin, Weiwen Liu, Yong Yu, Xing Sun and Weinan Zhang 26422

Meta-Moderator: Empowering Multi-Agent Debate with Meta-Cognition
Wentao Hu, Zhuoyue Wan, Jinhao Shen, Chen Jason Zhang, Xiao-Yong Wei and Li Qing . 26435

Using large language models to provide effective embeddings for native language identification
Roxane Auger, Olivier Boeffard and Jean-François Bonastre 26452

When Models Know but Can't Say: Relational Encoding Limits Linear Vulnerability Classification in Code LLMs
Rui Melo, Andre Catarino, Claudia Mamede, Corina S. Pasareanu and Rui Abreu 26461

Is Prosody Lost in Translation? Fine-Grained Cross-Lingual Prosody Similarity Across Languages
Haopeng Xie, Ismail Rasim Ulgen, Sofia Son, Berrak Sisman and Philipp Koehn 26498

CAGE: Coherence-Aware Graph Encoding for Retrieval-Augmented Generation
Tong Qi, Jingyu Wu, Youbing Yin, Spencer Hong, Daben Liu and Erin Babinsky 26512

ReContext: Recursive Evidence Replay as LLM Harness for Long-Context Reasoning
Yanjun Zhao, Ruizhong Qiu, Tianxin Wei, Yuanchen Bei, Zhining Liu, Lingjie Chen, Ismini Lourentzou, Hanghang Tong and Jingrui He 26531

EMMA: Event-Driven Multimodal Memory System with Multi-Agent Retrieval
Ruochen Zhao, Jiazheng Kang, Dongying Lin, Gezi Li and Ting Bai 26549

Bias Before the Model: Detecting and Removing Cognitive Bias Triggers in Financial Text
Triantafillos Papadopoulos, Venetis Charilaos, John Pavlopoulos, Ion Androutsopoulos and Sophia Ananiadou 26562

ChemVA: Bridging Visual and Semantic Gaps in Multimodal LLM Chemical Diagram Reasoning
Mingyang Rao, Kehua Feng, Zhihui Zhu, Jiangzhen Fu, Hao Yu, Keyan Ding and Huajun Chen 26588

Faithfulness Evaluation for Decoder-Only LLM Attributions with Controlled Retained Information
Xin Huang and Antoni B. Chan 26615

Med-Banana: Learning Quality-Controlled Medical Image Editing from Success-and-Failure Trajectories
Zhihui Chen, Qingyuan Lei, Kai He, Yanrui Du and Mengling Feng 26633

JFTA-Bench: Evaluate LLM's Ability of Tracking and Analyzing Malfunctions Using Fault Trees
Yuhui Wang, Zhixiong Yang, Ming Zhang, Shihan Dou, Zhiheng Xi, Di Liang, Enyu Zhou, Senjie Jin, Yujiong Shen, Dingwei Zhu, Yi Dong, Yujia Chen, Haibo Shi, Tao Gui, Qi Zhang and Xuanjing Huang 26653

Uncovering the Knowledge Boundaries and Mechanistic Knobof LLMs
Zihao Huang, Rui Mao, Yong Jiang and Erik Cambria 26679

Surgical Alignment in Knowledge Graph Training for Clinical Diagnosis with Large Language Models
Saksham Khatwani, He Cheng, Majid Afshar, Dmitriy Dligach and Yanjun Gao 26695

Extending LLM Context via Associative Recurrent Memory
Gleb Kuzmin, Ivan Rodkin, Aydar Bulatov, Yuri Kuratov, Lyudmila Rvanova, Mikhail Katkov, Ilia Sochenkov, Misha Tsodyks, Timothy Baldwin, Mikhail Burtsev and Artem Shelmanov 26710

MATE: Diagnosing Empathy Calibration Failures in Multi-Turn Human–LLM Interaction
Yeseon Hong, Junhyuk Choi, Minju Kim and Bugeun Kim 26743

BenGER: Benchmarking LLM Systems on Subsumption-Based Legal Reasoning in German Law
Sebastian Nagl, Ann-Kristin Mayrhofer, Martin Heidebach, Aleyna Koçak, Anne Zettelmeier, Elly Breu, Angelina Greiner, Sofija Milijas and Matthias Grabmair 26759

Structure Matters Semantic-Structural Dual-Path Arbitration for Named Entity Recognition with LLMs
Xinghong Hou, Yiqi Tong, Wei Chen, Wei Guo and Fuzhen Zhuang 26795

Skill-Conditioned Gated Self-Distillation for LLM Reasoning
Jiazhen Huang, Xiao Chen, Xiao Luo, Yong Dai, Senkang Hu and Yuzhi Zhao 26813

Retrieved But Not Reliable: A Survey on Attacks, and Defenses in Retrieval-Augmented Generation
Tran Ta Quang Minh, Cuong Dang, Tuc Nguyen, Khanh-Tung Tran, Minh Huynh Nguyen, Chau Nguyen To Trinh, Kien Le, Do Xuan Long, Jiahao Zhang, Fali Wang, Hoang D. Nguyen, Thanh Le and Suhang Wang 26832

PersonaSteer: Dynamic Activation Steering for Personalized Multi-turn LLM Conversations
Sijian Ren, Qingbin Zeng, Bingqiao Gu, Fengli Xu and Yong Li 26856

Dense Clinical Contrasts Enhance Medical Knowledge Updating in Large Language Models
Yangmin Huang, Shu Quan, He Geng, Xin Ye, Qianyun Du, Zhiyang He, Jiaxue Hu and Xiaodong Tao 26877

Perception Is the Bottleneck for Large Audio Reasoning Models
Ziyang Ma, Ruiyang Xu, Wenming Tu, Zhenghao Xing, Yuxuan Wang, Yunfei Chu, Jin Xu, Kai Yu, Junyang Lin, Eng Siong Chng and Xie Chen 26908

Towards Better Legal Reasoning LLMs: Signal Balancing and Reward Scheduling
Yinghao Hu, Qian Xiao, Yuting Huang, Kaisong Song, Yiquan Wu, Ang Li, Junjie Cao, Jun Lin, Kun Kuang and Fei Wu 26924

Harmless Yet Harmful: Neutral Prompting Attacks for Stealthy Hallucination Steering in Agent Skills
Chia-Yi Hsu, Chia-Mu Yu, Chun-Ying Huang and Jun Sakuma 26945

LLMs Can Predict Failure Risk, But Struggle to Predict Which Collaboration Protocol Pays Off: Cost-Aware Protocol Routing Across Reasoning Tasks
Chih-Hsuan Yang, Jingyan Jiang, Cheng-Hau Yang, Vikram Vasudevan, Huihuo Zheng, Venkatram Vishwanath and Rajeev Thakur 26995

LISAA: A Framework for Large Language Model Information Security Awareness Assessment
Ofir Cohen, Gil Ari Agmon, Rami Puzis and Asaf Shabtai 27017

Chain-of-Thought Hurts When the Answer Speaks for Itself
Shlok Shelat, Shrey Vimalbhai Salvi, Manas Gaur and Souvik Roy 27047

ThaiPaliBench: Benchmarking Large Language Models on Thai-Script Pali Understanding
Nobpharat Phetsri, Anon Saeoueng, Sarawoot Kongyoung and Waranrach Viriyavit 27072

Structured Strategy Injection for Data Analysis Agents
Zifu Tao, Guozhao Mo, Weixiang Zhou, Yaojie Lu, Hongyu Lin, Ben He, Xianpei Han and Le Sun 27100

Cross-Lingual Activation Steering for Visual-Prior Conflicts
Yufei Wang, Xiang Lorraine Li and Adriana Kovashka 27124

STAIF: A Stage-wise Optimization for Complex Instruction Following
Jian Hong, Chen Cheng, Quan Liu, Yuhao Chen and Enhong Chen 27136

RUBAS: Rubric-Based Reinforcement Learning for Agent Safety
Xian Qi Loye, Qinglin Su, Zhexin Zhang, Shiyao Cui, Qi Zhu, Fei Mi, Hongning Wang and Minlie Huang 27154

When Retrieval Partially Conflicts with Skills: Understanding Procedural Hallucination in Language Models
Guanqun Jiang, Pengfei Cao, Kang Liu and Jun Zhao 27176

ParaCalib: Semantically Calibrated Paralinguistic Modeling for Depression Detection
Yuxin Li, Yifei Li, Yi-Wen Chao, Xiangyu Zhang, Eng Siong Chng and Cuntai Guan 27197

GIFT: Reconciling Post-Training Objectives via Variational Finite-Temperature Gibbs Initialization
Zhengyang Zhao, Lu Ma, Yizhen Jiang, Xiaochen Ma, Zimo Meng, Chengyu Shen, Lexiang Tang, Haoze Sun, Peng Pei and Wentao Zhang 27215

Bridging Lexical Divergence: LLM-Assisted, Cost-Efficient, Zero-shot Scientific Entity Linking
Md Rasel Khondokar, Qiao Qiao, Farjana Sultana Samia, Nhat Le, Yuepei Li and Qi Li . . . 27231

ARMOR-PRM: Adversarially Robust Process Reward Models via Counterfactual Margins and Invariance Regularization
Hongyu Chen, Guangjia Chai, Haoxian Nan, Ruixue Wang and Peilin Jiang 27250

Toppling the Hierarchy in Byte-level Language Modeling
Lukas Edman and Alexander Fraser 27265

When Users Don't Ask: Benchmarking Context-Driven Memory Retrieval in Conversational Agents
Wen Yu Chang and Yun-Nung Chen 27273

Guided Permutation Scoring (GPS) : Option-Order Permutation for Multiple-Choice Question Answering
Md Shahrar Fatemi, Zhan Shi, Klaus Mueller and Tengfei Ma 27294

Reflecting What Matters: Same-Surgery References for Postoperative Clinical Note Generation
Seungmi Yu, Somin Song, Hyeryun Park and Jinwook Choi 27314

Do MLLMs Really Understand Space? A Mathematical Spatial Reasoning Evaluation
Shuo Lu, Jianjie Cheng, Yinuo Xu, Yongcan Yu, Lijun Sheng, Peijie Wang, Siru Jiang, Yongguan Hu, Lingxiao He, Meng Wang, Qianlong Xie, Xingxing Wang, Ran He and Jian Liang 27327

Agent Tools Orchestration Leaks More: Dataset, Benchmark, and Mitigation
Yuxuan Qiao, Dongqin Liu, Hongchang Yang, Wei Zhou and Songlin Hu 27344

CausalPrism: Component-Level Evaluation of Intervention-Based Causal Reasoning in Large Language Models
Jinyue Feng and Frank Rudzicz 27364

GAMS: Geospatial Knowledge Enhanced Agentic Framework for Urban Human Mobility Behavior Simulation
Yuwei Du, Jie Feng, Jian Yuan and Yong Li 27390

The Role of Entropy in Visual Grounding: Analysis and Optimization
Shuo Li, Jiajun Sun, Zhihao Zhang, Xiaoran Fan, Senjie Jin, Hui Li, Yuming Yang, Junjie Ye, Lixing Shen, Tao Ji, Tao Gui, Qi Zhang and Xuanjing Huang 27406

MIMIC-CTRL-2026: A Realistic Clinical Trial Dataset to Benchmark LLM performance
Noshin Nawal, Robert Margaryan, Moustafa Abdalla, Justin A. Ezekowitz and Mohamed Abdalla 27428

DuCEI: When to Introduce Metadata in Cross-Entity Interaction for Entity Disambiguation
Jinwoo Min, Roseline Nyange, Wonkee Lee, Hyun-Je Song and Seung-Hoon Na 27453

Beyond One-Hot Replay: Self-Distillation Reduces Target Variance in Continual Pre-Training
Peng Lu, Sijie Dong, Xiaoyun Feng, Xuanwei Hu and Xin Zhang 27473

When Linguistic and Internal Confidence Diverge in Large Language Models
Hefan Zhang, Bingquan Zhang, Ming Cheng, Saeed Hassanpour, Weicheng Ma and Soroush Vosoughi 27492

Prism: Persistent Faceted File Memory for Cost-Efficient Coding Agents
Ziyun Zhang, Yihao Liu, Yuzhe Zhang and Huaqian Cai 27518

EAGER: Enhancing Generative Event Extraction via Reinforcement Learning with Verifiable Rewards
Omar Adjali, Siting Liang, Omair Shahzad Bhatti and Daniel Sonntag 27532

TICD: Time-Aware Internal Contrastive Decoding for Enhancing Factuality in Diffusion Language Models
Zike Chen, Hongkun Dou, Mukang You, Yue Deng and Jiajun Li 27558

Evaluating multi-modal LLMs on historical book illustrations
Andrew Piper and Jiaming Jiang 27575

Same Ranking, Different Winner: How Scoring Targets Shape LLM Memory Benchmarks
Sugam Panthi and Rabab Abdelfattah 27590

Context Modeling for LLM-Based Behavioral Simulation
Dongyang Zhen, Ruize Qin, Ali Mansouri, Abdolmajid Erfani and Qingbin Cui 27607

Why are all LLMs Obsessed with Japanese Culture? On the Hidden Cultural and Regional Biases of LLMs
Joseba Fernandez de Landa, Carla Perez-Almendros and Jose Camacho-Collados 27628

AFROSCOPE: A Framework for Studying the Linguistic Landscape of Africa
Sang Yun Kwon, AbdelRahim A. Elmadany and Muhammad Abdul-Mageed 27651

VERGE : Formal Refinement and Guidance Engine for Verifiable LLM Reasoning
Vikash Singh, Darion Cassel, Nathaniel Weir, Nick Feng and Sam Bayless 27677

Causal Attribution Meets Impact-aware Verification: Debugging LLM-based Multi-Agent Systems via Static–Dynamic Synergy
Yawen Wang, Wenjie wu, Junjie Wang and Qing Wang 27703

Cross-Cultural Value Attribution in Large Vision-Language Models
Phillip Howard, Xin Su and Kathleen C. Fraser 27721

Query-Efficient Memory Injection Attacks on LLM Agents
Yicong Tan and Yang Zhang 27766

DramaSearch: Benchmarking Multimodal Searching Agents in Ultra-long TV Dramas
Yunning Cao, Jiannan Ge, Chuanbin Liu, Lingxi Xie, Xinyue Huo, Pengfei Chen, Kaiwen Duan, Hongtao Xie, Yongdong Zhang and Qi Tian 27798

Scalable Multi-Agent LLMs for Multi-Turn Conversation Bias Detection and Mitigation
Victor H. Vargas, Farsheed Haque, Depeng Xu and Xi Niu 27827

DARE: Dialectical Agentic Reasoning for Structured Knowledge Fact Checking
Yifei Li, Xiaohan Zheng, Wentao Qian and Liansheng Zhuang 27844

AEL: Evolving Agent Harness in Open-Ended Environments
Wujiang Xu, Jiaojiao Han, Minghao Guo, Kai Mei, Xi Zhu, Han Zhang and Dimitris N. Metaxas 27857

CourtGuard-PMPD: Zero-Shot Policy Adaptation in LLM Safety without Fine-Tuning using Prompt-Based Modular Policy Database
Umid Suleymanov, Rufiz Bayramov, Suad Gafarli, Taghi Mammadov, Seljan Musayeva and Murat Kantarcioglu 27886

Privacy-Preserving Heterogeneous Multi-LLM Federated Inference for Cognitive Diagnosis
Yagna Manasa Boyapati, Chong Yu, Tianyu Jiang and Justin Zhan 27913

Query Symbolically or Retrieve Semantically? A Dataset and Method for Semi-Structured Question Answering
Mateusz Czyżnikiewicz, Ryszard Tuora, Adam Kozakiewicz, Tomasz Ziętkiewicz, Mateusz Galiński, Michał Tomasz Godziszewski, Michal P. Karpowicz, Timothy Hospedales and Cristina Cornelio 27929

MedProb: Probing Internal Representations of Vision-Language Models for Medical Question Answering
Erfan Nourbakhsh, Ke Yang and Anthony Rios 27961

Attitude Analysis in Systemic Functional Linguistics: A New Corpus and Benchmark Task
Jorge Ortiz-Fuentes, Felipe Bravo-Marquez and Beatriz Quiroz 28000

Reduced Matrix Multiplication: Input-Adaptive Matrix-Product Reduction for LLM Inference
Zixuan Lan, Yanhong Li and Jiawei Zhou 28024

OISD: On-Policy Internal Self-Distillation of Language Models
Xinyu Liu, Darryl C. Jacob, Yang Zhou, Jindong Wang and Pan He 28049

A Pre-Generation Epistemic Routing Axis in Language Models
Steve Bakos, Heidar Davoudi and Aijun An 28066

CoBRA: Learning Tool-Use Boundaries via Counterfactual Margins
Wenhao Zou, Xianglong Liu, Wendong Bi, Hanjie Wang, Simin Zhao and Gong Zhi 28092

PEAM: Parametric Embodied Agent Memory for Continual Skill Learning via Internalization of Experience in Minecraft
Yuchen Guo, Junli Gong, Weicheng Wang, Hongmin Cai, Yiu-ming Cheung and Weifeng Su 28111

LEAF: Growing Trees Without Branching for Speech-Aware Large Language Model Post-Training
Argyrios Gerogiannis, Yekaterina Yegorova, Mark A. Hasegawa-Johnson and Venugopal Veeravalli 28125

Refusal Before Decoding: Detecting and Exploiting Refusal Signals in Intermediate LLM Activations
Matteo Gioele Collu, Riccardo Conte, Alberto Giaretta, Denis Kleyko, Mauro Conti, Matteo Zavatteri and Roberto Confalonieri 28141

Small Cues, Big Consequences: Learning Pivotal Cues for Multimodal Meme Classification
Akshit Sharma and Prashant W Patil 28158

Dropping the Anchor: Statistical Context Summarization for Distributed Systems via Pulsar Attention
Aryan Sood, Shantanu Acharya and Gaurav Kumar Nayak 28176

Folding Context into Executable States: Structured Strategy Folding for Long-Horizon Reasoning Agent
Zhengyu Chen, Luan Zhang, Zhaoyi Li, Jinluan Yang, Zheqi Lv, Teng Xiao, Wei Wang and Jingang Wang 28192

GROVE: Graph-based Routing Over Variant Experts for Adaptive Two-Stage Training of Reasoning Models
Yiming Luo, Chunlin Tian and Li Li 28210

MTGBench: A Multi-Turn Game Benchmark for Diverse LLM Capability Assessment
Alisa Alenicheva, Timur Zykov, Artsiom Saladukha and Vladislav Negodin 28225

LaMoC: Loss-Aware Modular Compression for LLMs
Mohanad Odema and Jacob Song 28280

DualGauge: Automated Joint Security-Functionality Benchmarking of Specification-Only Code Generation by LLMs and Coding Agents
Rupam Patir, Keyan Guo, Suvadra Barua, Jiawei Guo, Abhijeet Pathak, Dinesh Gudimetla, Hongxin Hu and Haipeng Cai 28295

Hate Is More Than What We See: Precedent-Guided Structured Reasoning for Hateful Video Detection
Junyi Chen, Qian Liu, Jing Sun and Yi Zhang 28320

MobileDreamer: Generative Sketch World Model for GUI Agent
Yilin Cao, Yufeng Zhong, Zhixiong Zeng, Liming Zheng, Jing Huang, Haibo Qiu, Peng Shi and Wenji Mao 28345

AU-Harness: An Open-Source Toolkit for Efficient and Unified Evaluation of AudioLLMs
Hoang H Nguyen, Sidharth Surapaneni, Akshay Kalkunte Suresh, Jash Mehta, Aman Tiwari, Oluwanifemi Bamgbose, Khyati Mahajan, Jash Shah, Shruthan Radhakrishna, Sathwik Tejaswi Madhusudhan, Vikas Yadav and Sai Rajeswar 28360

ClinConsensus: A Physician-Calibrated Benchmark for Evaluating Clinical Rubric Coverage in Chinese Medical LLMs
Xiang Zheng, Han Li, Wenjie Luo, Weiqi Zhai, Yiyuan Li, Chuanmiao Yan, Xue Yang, Kailuan Wu, Ruiyi Xu, Tianyun Lu, Tianyi Tang, Yubo Ma, Kexin Yang, Sen Yang, Dayiheng Liu, Lin Qu, Bing Zhao and HU Wei 28381

FairFund-Bench: Evaluating Distributive Bias in LLM Resource Allocation
Martin Lukk 28406

From Process Loss to Assembly Bonus: Human-Grounded Diagnosis of Multi-Agent LLM Collaboration
Ala N. Tak, Teruhisa Misu, Kumar Akash, Zhaobo Zheng, Kevin Joo and Jonathan Gratch 28425

The Scientific Contribution Graph: Automated Literature-based Technological Roadmapping at Scale
Peter Jansen 28448

CyberChainBench: Can AI Agents Secure Smart Contracts Against Real-World On-Chain Vulnerabilities?
Jintao Huang, Fengqing Jiang, Radha Poovendran and Zhiqiang Lin 28476

DRFLOW: A Deep Research Benchmark for Personalized Workflow Prediction
Md Tawkat Islam Khondaker, Raymond Li, Muhammad Abdul-Mageed, Laks V. S. Lakshmanan and Issam H. Laradji 28492

CodeEvolve: an open-source evolutionary framework for algorithmic discovery and optimization
Henrique Assumpção, Diego Leonardo Braga Ferreira, Leandro Augusto Lacerda Campos and Fabricio Murai 28559

The Geometry of Ignorance: Detecting Suppressed Knowledge in LLMs through Activation Fragmentation
Adar Peleg, Dvir Alsheich, Tomer Ashuach and Avi Mendelson 28582

DART: Draft-Agreement Routing for Training-Free Adaptive Thinking Budgets in Hybrid Reasoning Models
Jungseob Lee, Seongtae Hong, Seungjun Lee, Jaehyung Seo, Junyoung Son, Sugyeong Eo, Chanjun Park, Hyeongju Park, Hyeonseok Moon and Heuiseok Lim 28609

Translation Circuit Discovery and Targeted Optimization for Low-Resource Multilingual Translation
Atai Waris Khan, Abdullah Mazhar, Raghu Kumar and Md Shad Akhtar 28626

Empirically Grounded Behavioral Adaptivity for LLM-Based Virtual Patients in Psychotherapy Training
Angela Chen, Siwei Jin, Catherine Bao, Canwen Wang, Yi-Chia Wang, Robert E. Kraut, Tongshuang Wu and Haiyi Zhu 28644

Latent-Level Uncertainty Resolution for Stable EEG-to-Text Decoding under Cross-Subject Variability
Sejin Park, Jedong Lee, Yunhui So and Byung Hyung Kim 28666

Towards Automatic Evolution Tree Generation from Citation Graphs
Zexing Zhao, Yuntong Hu and Liang Zhao 28680

AsyncWebRL: Efficient Asynchronous Reinforcement Learning for Multi-Step Visual Web Agents
Hao Bai, Rui Yang, Chenlu Ye, Spencer Whitehead, Aviral Kumar and Tong Zhang 28704

Learning to Reason by Analogy via Retrieval-Augmented Reinforcement Fine-Tuning
Zilin Xiao, Qi Ma, Chun-cheng Jason Chen, Xintao Chen, Avinash Atreya, Hanjie Chen and Vicente Ordonez 28723

Query Timing Produces Opposite Positional Biases Between LLMs and Humans
Jasin Cekinmez, Addison J. Wu and Thomas L. Griffiths 28741

Measuring Cultural Alignment Beyond the Average: A Framework for Evaluating Maternal-Health LLM Interactions in Indian Contexts
Umaira Izhar, Gunjan Arora and Pushpendra Singh 28759

CoSTA: Cost-Sensitive Toolpath Agent for Multi-turn Image Editing*
Advait Gupta, NandaKiran Velaga, Dang Nguyen and Tianyi Zhou 28779

CLAMP: Constrained Decoding for Vision-Language Embodied Planning
Tianyi Ma and Parisa Kordjamshidi 28815

MCPGen: Benchmarking LLMs on Executable MCP Workflow Development
Yingxuan Yang, Jiaqi Liu, Lirui Guan, Jiaye Gao, Weiwen Liu, Weinan Zhang and Ying Wen 28846

RW2Unsafe: Towards More Accurate Real-World Unsafe Generation Evaluation for T2I Models
Ye Leng, Junjie Chu, Mingjie Li, Michael Backes, Yun Shen and Yang Zhang 28860

StructuralCost: A controlled reading time dataset for modeling human sentence processing difficulty
Nina Nusbaumer, Iria de-Dios-Flores, Corentin Bel, Christophe Pallier, Guillaume Wisniewski and Benoit Crabbé 28874

Detector-Evasive LLM Paraphrasing via Constrained Policy Optimization
Mingyi Wang, Zhuoer Shen, Yuheng Bu and Shaofeng Zou 28889

CROCODIL: Cross-Model Code Editing with LLMs
Linghan Zhong, Aditya Thimmaiah, Jayanth Srinivasa, Milos Gligoric and Junyi Jessy Li . 28915

Analyzing the Effect of Media Takeovers on their Political Positioning
Lubin Longuépée, Akram Elbouanani and Adrian Popescu 28934

Vendi Clustering for Diversity-Preserving Topic Reduction
Sean Kim and Adji Bousso Dieng 28964

Towards Fine-grained and Robust Detection of Human–AI Co-authorship
Ziruo Zhao, Jie Fu, Wendy Hui Wang and Zining Zhu 28981

EvolvingWorld: An Open-Schema Framework for Co-Evolving Role-Play Agents and World Model in Interactive Literary World
Qing Zong, Yue Guo, Mengxin Yang, Steven Y. Guo and Yangqiu Song 28999

Diffusion-Proof: Recipe for Formal Theorem Proving Beyond Auto-Regressive Generation
Ruida Wang, Rui Pan, Pengcheng Wang, Shizhe Diao and Tong Zhang 29050

Do VLMs Read or Rewrite? On Transcription Faithfulness in Vision-Language Models
Gwang Gook Lee, Kenan E. Ak, Jay Mohta, Yan Xu and Dimitrios Dimitriadis 29077

CroCo: Cross-Lingual Contrastive Preference Tuning on Self-Generations
Mike Zhang, Ali Basirat and Desmond Elliott 29093

OmniLayout: A Schematic-Coupled Multimodal Benchmark for Geometric Reasoning in PCB Layout
Taiting Lu, Kaiyuan Lin, Mingjia Wang, Haolin Ye, Runze Liu, Yuxin Tian, Vahe Melkonyan, Haoyu Wang, Muchuan Wang, Chufan Hong, Yifan Yang, Sung-Liang Chen, Yi-Chao Chen, Yincheng Jin and Mahanth Gowda 29121

Thinking While Speaking: Inference-Time Knowledge Transfer for Responsive and Intelligent Conversational Voice Agents
Vidya Srinivas, Zachary Englhardt, Vikram Iyer and Shwetak Patel 29148

Disappearing Ink: Obfuscation Breaks N-gram Code Watermarks in Theory and Practice
Gehao Zhang, Mingzhe Li, Eugene Bagdasarian, Shiqing Ma and Juan Zhai 29188

URBAN-PROTOS: Discrete Semantic Prototypes for Cross-City Region Representation Learning
Zihang Xu, Xinshun Feng, Haozhi Xie, Yi Qiao, Liangzhe Han and Leilei Sun 29220

Deference with attributions in LLM Cascades
Lize Pirenne, Gaspard Lambrechts, Norman Marlier, Maxence De La Brassinne Bonardeaux, Gilles Louppe and Damien Ernst 29236

Interpretable Stylistic Variation in Human and LLM Writing Across Genres, Models, and Decoding Strategies
Swati Rallapalli, Shannon K. Gallagher, Ronald Yurko, Tyler Brooks, Chuck Loughin, Michele Sezgin and Violet Turri 29243

HypoKG: Evidence-Disciplined Biomedical Hypothesis Generation Beyond Endpoint Knowledge
Dominic Okonkwo, Adetayo O. Okunoye and I. Budak Arpinar 29262

Prioritizing the Best: Incentivizing Reliable Multimodal Reasoning by Rewarding Beyond Answer Correctness
Mengzhao Jia, Zhihan Zhang and Meng Jiang 29279

Why TaylorMLP Fails? Revisiting LLM Weight Protection Schemes with Adversarial Games
Bintao Tang, Xin Yang, Letian Li, Yu Meng, Kwok Wai Lui, Yuhao Wang, Zimo Ji, Terry Jingchen Zhang and Wenyuan Jiang 29296

Emulate or Estimate? The Divergent Strengths of Base and Post-Trained Language Models for Opinion Simulation
Seth Grief-Albert, Jessica Bo, Difan Jiao and Ashton Anderson 29311

Characterizing and assessing robust conversational interaction: Amendments in order specification
Simon Benigeri and Larry Birnbaum 29321

Tuning Just Enough: Lightweight Backdoor Attacks on Multi-Encoder Diffusion Models
Ziyuan Chen, Yujin Jeong, Tobias Braun and Anna Rohrbach 29361

Massive Spikes in LLMs are Bias Vectors: Mechanistic Uncovering and Spike-Free Quantization
Yung-Chin Chen, Chung Peng Lee, Ze-Wei Liou and Naveen Verma 29380

Translating the Untranslatable: An Operationalizable Ontology for Untranslatability
Jacob Bremerman, Brihi Joshi, Hirona Jacqueline Arai, Xiang Ren and Jonathan May 29405

Generating Clinical Vignettes that Preserve Cognitive Formulations
Amit Oren, Nimrod Hertz-Palmor, Dean Ariel and Guy Laban 29423

MechPunBench: A Bilingual Mechanism-Centered Benchmark for Pun Understanding and Generation
Tianwei Zhang, Zihan Gu, Jiayu Wang, Tianyi Ma, Jiacheng Lu, Jinyu Li and Hua Zhang . 29473

Adaptive Cost-Efficient Evaluation for Reliable Patent Claim Generation
Yongmin Yoo, Qiongkai Xu and Longbing Cao 29501

Large Reasoning Models Struggle to Transfer Parametric Knowledge Across Scripts
Lucas Bandarkar, Alan Ansell and Trevor Cohn 29519

Tracing the Latent Threads: A Mechanistic Study of How LLMs Represent and Operationalize Race and Ethnicity Cues
Shiyue Hu, Ruizhe Li and Yanjun Gao 29530

CROWDMATH: A Dataset of Crowdsourced Mathematical Research Discussions
Sherin Muckatira, Jesse Geneson, Slava Gerovitch, Pavel Etingof, Mikhail Gronas and Anna Rumshisky 29546

Latent-LoRA: Compact Latent-Space Adapters with Gradient-Free Routing for Continual Learning
Reza Rahimi Azghan, Gautham Krishna Gudur, Pavan K. Turaga, Giulia Pedrielli and Hassan Ghasemzadeh 29563

BalanceRAG: Joint Risk Calibration for Cascaded Retrieval-Augmented Generation
Zijun Jia, Yuanchang Ye, Sen Jia, Yiyao Qian, Haoning Wang, Baojie Chen, Diyin Tang, Jinsong Yu and Zhiyuan Wang 29581

Fork-Merge Decoding: Enhancing Multimodal Understanding in Audio-Visual Large Language Models
Chaeyoung Jung, Youngjoon Jang, Jongmin Choi and Joon Son Chung 29610

From Static Templates to Dynamic Runtime Graphs: A Survey of Workflow Optimization for LLM Agents
Ling Yue, Kushal Raj Bhandari, Ching-Yun Ko, Dhaval C Patel, Shuxin Lin, Nianjun Zhou, Jianxi Gao, Pin-Yu Chen and Shaowu Pan 29630

Towards Structured Mechanistic Reasoning in Single-cell Prediction
Yunhui Jang, Lu Zhu, Jake Fawkes, Alisandra Kaye Denton, Dominique Beaini and Emmanuel Noutahi 29659

GenoMAS: Guided Multi-Agent Programming for Scientific Discovery from Genomic Data
Haoyang Liu, Yijiang Li and Haohan Wang 29695

Think You Have Solved Commonsense Reasoning? Try HellaSwagUltra
Wenhan Han, Yifan Zhang, Binbin Liu, Mykola Pechenizkiy, Yin Zheng and Meng Fang . 29739

Disentangling Representation Evolution in Transformers through Directional Decomposition
Shwai He, Haichao Zhang and Shen Yan 29759

CANE is a Format: Rethinking Output Formats for Few-Shot LLM-Based Named Entity Recognition
Joseph Cornelius, Kei K Ito and Fabio Rinaldi 29773

Understanding the Role of Prompt Template in Knowledge Distillation for Safety Alignment
Anjila Budathoki, Manish Dhakal, Benjamin M. Ampel and Yi Ding 29801

Listen to the Latents: Self-Correcting Speech Recognition in Large Audio Language Models Through Hidden-State Interactions
Chan-Jan Hsu, Jaeyeon Kim, Chao-Han Huck Yang, Shinji Watanabe, Hung-yi Lee and Carlos Busso 29818

COGTRL: Training LLMs for Scientific Discovery Assistance using Cognitive Traces via Reinforcement Learning
Shrinidhi Kumbhar, Santosh Mashetty, Divij Handa, Kevin Coutinho, Siddharth Sambhaji Ghule and Chitta Baral 29842

MLOps-Bench: Benchmarking AI Agents on Production ML Engineering in Real Repositories
Quoc Co Tran, Hitesh Laxmichand Patel, Diwakar Mahajan, Kshitij Bakliwal, Avi Sil and Katrin Kirchhoff 29888

Register Shifts Break LLM Safety: A Bengali Benchmark with Culturally Grounded Harms
Naymul Islam, Nusrat Jahan Lia, Shubhashis Roy Dipta, Sabik Bin Sultan and Abdullah Khan Zehady 29904

CloRe: Closed-Loop Reasoning via Intermediate-State Maintenance for Multi-Hop Question Answering
Yingxin Pei, Yuantao Han, Ruiting Li, Binquan Ji, Haibo Luo, Yifei Lu and Feiliang Ren . 29924

Compliance vs. Sensibility: On the Reasoning Controllability in Large Language Models
Xingwei Tan, Marco Valentino, Mahmud Elahi Akhter, Yuxiang Zhou, Maria Liakata and Nikolaos Aletras 29954

GinSign: Grounding Natural Language Into System Signatures for Temporal Logic Translation
William H English, Chase Walker, Dominic Simon and Rickard Ewetz 29970

Tool-R0: Self-Evolving LLM Agents for Tool-Learning from Zero Data
Emre Can Acikgoz, Cheng Qian, Jonas Hübotter, Heng Ji, Dilek Hakkani-Tür and Gokhan Tur 29989

Learning Composable Chains-of-Thought
Fangcong Yin, Zeyu Leo Liu, Liu Leqi, Xi Ye and Greg Durrett 30018

Randomized YaRN Improves Length Generalization for Long-Context Reasoning
Manas Mehta, Fangcong Yin and Greg Durrett 30045

CITAR-NER-Br: A Unified NER Dataset for Fine-Grained Legal Citation Extraction in Brazilian Portuguese
Arthur Elwing Torres, Altigran da Silva, Hidelberg Oliveira Albuquerque, Ellen Polliana Ramos Souza Pereira, Héldon José Oliveira Albuquerque, Nádia Félix Felipe Da Silva, Márcio De Souza Dias, Douglas Vitório, Andre Suslik Spritzer, Laura Pereira de Gouveia, Rafael Pereira de Gouveia and Breno Gabriel De Melo Lima 30056

Composition Collapse: Stable Factual Knowledge Does Not Imply Compositional Reasoning
Zhe Yu, Wenpeng Xing, Yunzhao Wei, Jie Chen, Hongzhi Wang, Xuyang Teng and Meng Han 30070

Step-by-Step CCG Parsing with Large Language Models
Tu-Anh Tran, Namgi Han, Taiga Ishii and Yusuke Miyao 30093

TextNCA: Neural Cellular Automata for Language Modeling via Hierarchical Local Attention
Avni Mittal, Avinash Anand, Ashutosh Kumar, Dikshant Kukreja, Kritarth Prasad, Sushane Dulloo, Erik Cambria, Timothy Liu, Zhengkui Wang and Rajiv Ratn Shah 30116

IDEEA: training-free Input-Dependent stEEring via Activation cluster matching
Zheng Wang, Muchen Li, Renjie Liao and Yan Leng 30142

Long-Context Demonstration Selection Using State Space Models
Ziniu Zhang, Zhenshuo Zhang, Ruoxuan Xiong, Gene Cooperman and Hongyang R. Zhang30161

Agentic Uncertainty Quantification
Jiaxin Zhang, Prafulla Kumar Choubey, Kung-Hsiang Huang, Caiming Xiong and Chien-Sheng Wu 30179

Retrieval-Augmented Foundation Models for Matched Molecular Pair Transformations to Recapitulate Medicinal Chemistry Intuition
Bo Pan, Yang Qiao, Peter Zhiping Zhang, Hao-Wei Pang, Alex Zhu, Xiang Yu, Liying Zhang and Liang Zhao 30210

Thinking with Visual Grounding
Junkai Zhang, Yihe Deng, Kai-Wei Chang and Wei Wang 30224

RadJEPA: Radiology Encoder for Chest X-Rays via Joint Embedding Predictive Architecture
Anas Anwarul Haq Khan, Mariam Husain, Pratik Jalan and Kshitij Sharad Jadhav 30238

Many-Tier Instruction Hierarchy in LLM Agents
Jingyu Zhang, Tianjian Li, William Jurayj, Hongyuan Zhan, Benjamin Van Durme and Daniel Khashabi 30260

SyPS: Measuring Sycophancy Prompt Sensitivity in Large Language Models
Lijia Huang, Yao Fu and Sihao Ren 30283

MMGist: A Comprehensive Multimodal Benchmark for 2027
Wenzhen Yuan, Jiacheng Ruan, Wutao Xiong, Chengping Zhao, Ting Liu and Yuzhuo fu . 30298

When Reasoning Meets Structure: Understanding Reasoning Strategies for Structured Analysis
Xin Qiu and Jialingan 30315

WeClawArena: An Auditable Sandbox and Benchmark for Cross-User Agents Collaboration and Security in Human-Centered Agent Networks
Prince Zizhuang Wang, Aojie Yuan, Haiyue Zhang, Xiyang Hu, Yue Zhao and Shuli Jiang 30350

ADRD-Bench: An LLM Benchmark for Alzheimer's Disease and Related Dementias
Guangxin Zhao, Jiahao Zheng, Malaz Boustani, Jaroslaw Nabrzyski, Yiyu Shi, Meng Jiang and Zhi Zheng 30378

FlexEE: Self-Speculative and KV-Compatible Early Exiting for Offloading-Aware LLM Inference
Qihu Xie, Ziwei Li and Yi Kang 30396

Agent Skill Evaluation and Evolution: Frameworks and Benchmarks
Kexin Ding, Yang Zhou, Can Jin, Feng Tong, Mu Zhou and Dimitris N. Metaxas 30412

Intent-Driven Situation Tracking for User-Centric Multi-Turn Agents
Meiling Tao, Yiling Tao and Peng Wang 30432

Improving Language Identification for Code-Switched Utterances with Integer Linear Programming
Joanna Radoła, Josep Crego and François Yvon 30450

Analyzing Traditional and Neural Approaches to Multilingual Readability Assessment
Joshua Wong and Chris Tanner 30469

Domain Mixture Design via Log-Likelihood Differences for Aligning Language Models with a Target Model
Ryo Kishino, Riku Shiomi, Hiroaki Yamagiwa, Momose Oyama and Hidetoshi Shimodaira 30483

FinRAQ: Financial News Reasoning through Importance Ranking and Socratic Questioning
Zee Hen Tang and Mi-Yen Yeh 30500

IGTF: Intent-Guided, Game-Inspired Structured Intent Calibration for LLM-Assisted Fake News Detection
Kun Xu, Zongyuan Mou, Yuan Rao, Yufei Xin and Tingting Xin 30548

ActiveMedAgent: Cost-Aware Trajectory Learning for Multimodal Medical Diagnosis
Weiwei Ma, Xiaobing Yu, Peijie Qiu, Jin Yang, Zhaoqi An, Xuanzhao Dong, Xiaoqi Zhao and Xiaofeng Liu 30576

Attention-Steered Vision-Language Models for Sign Language Translation
Meibo Hu, Guohao Sun, Annemarie D. Ross, Sheng Li and Zhiqiang Tao 30596

AblationBench: Evaluating Automated Planning of Ablations in Empirical AI Research
Talor Abramovich and Gal Chechik 30612

MP-Bench: Evaluating Voice Agents as a Multiparty Conversation Participant
Yi-Jen Shih, Shih-Yun Shan Kuan, Guan-Ting Lin, Kai-Wei Chang, Siddhant Arora, Shu-wen Yang, Abdelrahman Mohamed, Shinji Watanabe, Hung-yi Lee and David Harwath 30652

From Citations to Contributions: LLM-Assisted Credit Scoring of Research Articles
Sana Ebrahimi, Suraj Shetiya and Abolfazl Asudeh 30672

AIM: Anchor Identity Features, Then Match for Multimodal Large Language Model Unlearning
Wonjun Lee, Jaehyuk Jang, Kangwook Ko, Hee-Seon Kim and Changick Kim 30699

LLM-assisted Data Quality Improvement: SRL as a Case Study
Lijie Wang, Yahui Liu, Yue Zhang and Zhenghua Li 30723

When Reflection Helps: Utility-Guided Advantage Rescaling for Efficient Reasoning
Jungyu Lee, Kunhui Lee, Dongwook Min, Jeonhwi Kang and Seung-Hoon Na 30739

Identifying Temporal Features within Transcoders for Time Sensitive Factual Recall
Sanjay Govindan, Maurice Pagnucco and Yang Song 30761

Computational surprisal is not enough: Discourse prominence modulates human eye-movement patterns but not GPT expectations in Korean
Nayoung Kwon, Gyu-Ho Shin and Seongmin Mun 30783

Am I More Pointwise or Pairwise? Revealing Position Bias in Rubric-Based LLM-as-a-Judge
Yuzheng Xu, Tosho Hirasawa, Tadashi Kozuno and Yoshitaka Ushiku 30797

Perceive-Then-Adjudicate: A Multi-Agent Framework for Resolving Knowledge-Evidence Conflicts in LLM Fact-Checking
Yongcheng Zhang, Xin Zhang, Fei Li, Chong Teng, Jiang-Ming Yang, Donghong Ji and Zhuang Li 30813

Reward Engineering for Software Tasks: A Survey of Reinforcement Learning Approaches
Md Rayhanul Masud, Azmine Toushik Wasi, Salman Rahman and Md Rizwan Parvez 30834

Divide, Conquer, and Distill: Unified Sample Selection and LLM Distillation for Noisy Long-Tailed Learning
Bo Yuan, Yulin Chen, Jie Cao and Mei Gao 30847

EviDocVLM: Evidence-Aligned Reinforcement Learning for Multi-Page Document Question Answering
Anran Liu, Zongsheng Cao, Jun Xie, Feng Chen and Zigan Wang 30865

Intrinsic Attention as Navigator: A Non-Generative Retrieval Method for Multi-Hop Queries
Linwei Song, Weijie Zhang, WeiDi Chen, Huixia Lai, Shi Zhang and Ruliang Xiao 30876

HAP: Head-Adaptive Visual Token Pruning via Cross-Modal Alignment
Yuanhao Sun, Huawei Ji, Yuan Jin, Cheng Deng, Luoyi Fu and Xinbing Wang 30893

Seeing Through Conflicts: Improving Instruction Hierarchy Alignment in Vision-Language Models
Nicholas Sansoterra, Zishuo Zheng and Sachin Kumar 30912

MindUntangler: Resolving Cognitive Entanglements via Multi-Agent Topological Routing in Psychological Counseling
Xiaoyu Liu, Yufeng Diao, Jingjie Zeng, Shiqi Li, Hao Zhang and Liang Yang 30928

MASkills: Continual Skills Optimization for Multi-Agent LLM Systems
Huaiyuan Yao, Xiaoou Liu, Charles Fleming, Tianlong Chen and Hua Wei 30947

Can Multimodal Large Language Models Generate and Detect Multimodal Social Media Fake News?
Jiyao Yang, Yang Liu, Zhenyue Qin, Qingyu Chen and Xiuzhen Zhang 30958

Beyond Prompts: Measuring and Optimizing LLM Tool-Agent Harnesses
Cen Zhao, Haibo Ruan, Wenjie Chen, Pei-fen Tu, Usman Abbasi and Joel Hesch 30973

What Gets Unmasked First? Trajectory Analysis of Diffusion Models for Graph-to-Text Generation
Qing Wang, Jacob Devasier and Chengkai Li 30992

SVE-Loom: Improving Session-Level Reliability for Natural-Language-to-SVE Code Generation with Bounded Feedback
Haoyuan Gui, Hongjiao Sheng, Yifan Zhang and Huiyuan Li 31013

Notes2Skills: From Lab Notebooks to Certainty-Aware Scientific Agent Skills
Shi Liu, Jiayao Chen, Chengwei Qin, Yanqing HU, Jufan Zhang and Linyi Yang 31030

What Does Layer-Importance Reveal About Transformers and State-Space Models?
Istabrak Abbes, Nizar Islah, Irina Rish and Sarath Chandar 31048

SHARD: Safe and Helpful Alignment via Self-Reframing Distillation
Viswonathan Manoranjan, Amogh Gupta, Anvesh Rao Vijjini, Thomas Hofweber and Snigdha Chaturvedi 31076

Empowering LLM Agents with Geospatial Awareness: Toward Grounded Reasoning for Wildfire Response
Yiheng Chen, Lingyao Li, Zihui Ma, Qikai Hu, Yilun Zhu, Min Deng and Runlong Yu . . . 31104

RoboTrustBench: Benchmarking the Trustworthiness of Video World Models for Robotic Manipulation
Huiqiong Li, Jiayu Wang, Zhiting Mei, Anirudha Majumdar, Jingjing Chen and Bin Zhu . . 31121

FedLoRAGuard: Federated Dynamic Graph Neural Networks with Differential-Privacy Certificates for LoRA Adapter Integrity Verification
Roger Nick Anaedevha, Alexander Trofimov and Borodachev Yuri Vladimirovich 31148

When Models Hear What They Expect: Diagnosing Prosodic Heuristics in Multimodal Sarcasm Detection
Yongjian Chen, Pengfei Wei, Yiqun Sun, Zhu Li and Lawrence B. Hsieh 31162

BudgetThinker: Empowering Budget-aware LLM Reasoning with Control Tokens
Hao Wen, Xinrui Wu, Yi Sun, Feifei Zhang, Liye Chen, Jie Wang, Yifan Su, Yunxin Liu, Ya-Qin Zhang and Yuanchun Li 31182

Does Thinking Help Fairness? Reasoning Tokens Resolve Some Biases but Create More
Deng Pan, Joe Germino, Yihong Ma, Elizabeth M. Daly, Nuno Moniz, Ting Hua and Nitesh V. Chawla 31206

AdaCultureSafe: Adaptive Cultural Safety Grounded by Cultural Knowledge in Large Language Models
Hankun Kang, Di Lin, Zhirong Liao, Pengfei Bai, Xinyi Zeng, Jiawei Jiang, Yuanyuan Zhu, Dexi Liu and Tieyun Qian 31223

SDRAG: Sub-question Dependency-Aware Retrieval-Augmented Generation for Multi-hop Question Answering
Mengze Wang, Huan Liu, Yixiang Dong, Lingling Zhang, Bo Hu and Zhuohang Li 31248

ParaThinker: Native Parallel Reasoning for Scaling LLM Test-Time Compute
Hao Wen, Yifan Su, Feifei Zhang, Xinrui Wu, Yi Sun, Yunxin Liu, Yunhao Liu, Ya-Qin Zhang and Yuanchun Li 31265

Orchestrating Hidden-Intermediate Pruning-and-Distill for MoEs Slimming
Yifu Ding, Yongcheng Jing, Chengyu Wang, Jun Huang and Dacheng Tao 31286

Towards Reliable Truth-Aligned Uncertainty Estimation in Large Language Models
Ponhvoan Srey, Quang Minh Nguyen, Xiaobao Wu and Anh Tuan Luu 31306

ProEvent: An Event-centric Benchmark for Proactive Agents
Guanzhen Li, Liangming Pan and Leye Wang 31322

FCPRAG: Fusion-Controller Parametric Retrieval-Augmented Generation for Stable Multi-Passage LoRA Injection
Jinchang Zhu, Jindong Li, YI Ding, Nie Xiaojian, Rong Fu, Shuangyong Song, Haowei He and Menglin Yang 31337

Critical Positions in LLM Reasoning: A Metacognitive Lens on Intervention-Sensitive Locations
Wendi Cui, Anni Huang, Wei Wang, Damien Lopez and Sirui Bi 31357

TopoEval: A Comprehensive Benchmark for Topological Reasoning in Foundation Models
Yinzhu Cheng, Yihan Meng, Haihua Xie and Mingming Sun 31391

Old Memories Die Hard: Unveiling Shallow Forgetting and Ripple Effects in Privacy Unlearning
Xiaoyi Chen, Haoyuan Wang, Siyuan Tang, Sijia Liu, Liya Su, XiaoFeng Wang and Haixu Tang 31422

ReCite: Agentic Reasoning for Faithful Citation
Yuyang Huang, Bobo Li, JiaJia Song, Yuzhe Ding, Chong Teng, Fei Li and Donghong Ji . 31441

Silence over Error: Risk-Averse Navigation Landmark Extraction via Progressive MLLM Alignment
Xin Yang, Ziqi Cao, Changning Shao, Junhao Lin, Fanxiang Zeng and Yue Liu 31455

HarmProfile: Characterizing Harmful Distributions in Frontier LLMs
Zhouyuan Ma, Yutao Wu, Hanxun Huang, Xiang Zheng, Xiao Liu, Yixin Cao, Zuxuan Wu, Xingjun Ma and Yu-Gang Jiang 31477

ClinTraceBench: Source-Verifiable Longitudinal Clinical Reasoning over EHR-Derived Dialogues
Huimin Wang, Zhengyi Zhao and Yutian Zhao 31524

ACE: Affective Continuous Editing via Object–Atmospheric Control for Fine-Grained Image Editing
Hao Li, Kai Shu, Gang Yang, Jingyang Jia, Xun Chen and Aiping Liu 31550

Factual Decoding via Head Coordination Graphs in Large Language Models
Shike Li, Xiaokai Wang, Xin Tong, Xiaofeng Liu and Hu Zhang 31569

MIS-Bench: Benchmarking Multimodal LLMs for Psychotherapeutic Interpersonal Skills Assessment
Yuhan Lu, Yi Yao, Hua Shen, Katie Aafjes-van Doorn and Zhaonan Wang 31582

MegaCode: A Repository-Level Dataset for Long-Context Code Reasoning
Wenyang Luo, Shuang Sun, Yiwen Hu, Fei Bai, Yanzipeng Gao, Shuo Wang, Hongyu Lu, Xin Zhao and Ji-Rong Wen 31608

CAMFT: Conflict-Aware Mergeable Fine-Tuning for Large Language Models
Jingang Zhou, Haiyang Guo, Yuan Ma, Han Zhu and Xu-Yao Zhang 31627

Avoidance Learning of Failure Modes Produces Robust Tool Planners
Hadi Askari, Japinder Singh, Tenghao Huang, Dongwon Jung, Anshuman Chhabra and Muhao Chen 31641

REIGN: Refurbished Embeddings with Integrated Guidance Networks for Efficient Context-Length Scaling
Devrim Çavuşoğlu and Emre Akbas 31659

When 2D Tasks Meet 1D Serialization: On Serialization Friction in Structured Tasks
Chung-Hsiang Lo, Lu Li, Diji Yang, Tianyu Zhang, Yunkai Zhang, Yoshua Bengio and Yi Zhang 31678

Benchmarking Argument Identification and Ellipsis Recovery in Chinese Comparative Constructions
Wenjing Xu and Hongzhi Xu 31694

OPE: Overcoming Information Saturation in Parallel Thinking via Outline-Guided Path Exploration
Qi Guo, Jianing Wang, Deyang Kong, Xiangyu Xi, Jianfei Zhang, Yi Lu, Jingang Wang, Wei Wang, Shikun Zhang and Wei Ye 31708

Beyond Similarity: Coverage-Aware Prompt Selection for Time Series Forecasting with LLMs
Daeun Ji, Minkyoung Kim, Dongkuk Kim, Yohan Lee, Beomsoo Kim and Beakcheol Jang31723

BlockBatch: Multi-Scale Consensus Decoding for Efficient Diffusion Language Model Inference
Xiaoyou Wu, Cheng-Jhih Shih, Binfei Ji, Yong Liu and Yingyan Celine Lin 31747

AMDKernelVault: Large-Scale Datasets and Agentic Training for AMD GPU Kernel Optimization
Ji Liu, Saptarshi Majumder, Yiqing Huang, Wenwen Ouyang, Umang Pandey, Zeping Li, Chushi Chen, Zihao An, Puyuan Yang, Zekai Li, Sina Rafati, Ziqiong Liu, Pratik Prabhanjan Brahma, Dong Li, Zicheng Liu, Sharon Zhou and Emad Barsoum 31775

KGFlow: Enhancing Large Language Model Reasoning with Multi-Graph Alignment for Knowledge-Intensive Medical Question Answering
Tiantian Zhu, Xiangdong Kong, Xizhao Wang, Wing W. Y. Ng, Yuheng Jia, Jun Zhang and Zexuan Zhu 31799

CWF: A Collaborative Writing Framework for Personalized and Reliable Long-Form Popular Science Writing
Ruibiao Fu, Di Tang, Yunlong Yang, Ran Wang, Sicheng Lu, Peixuan Wu, Xiaoyu Fan, Jiacheng Ma, HaoZhe Luo and Yang Xiao 31820

Compositional Generalization with Outcome-Level Optimization
Xiyan Fu and Wei Liu 31845

From Perception to Simulation: Unveiling and Empowering Mental Spatial Dynamics in MLLMs
Jiachun Li, Zhuoran Jin, Kang Liu and Jun Zhao 31858

SteganoBackdoor: Evading Data-Poisoning Defenses via Steganographic Backdoors
Eric Xue, Ruiyi Zhang and Pengtao Xie 31886

Vendi-RAG: Adaptively Trading-Off Diversity And Quality Significantly Improves Retrieval Augmented Generation With LLMs
Mohammad Reza Rezaei and Adji Bousso Dieng 31909

Which Language Should a Skeleton Speak? Language Choices in Multilingual Reasoning
HyeonSeok Lim, SeungWoo Song, Inho Won, Hoyun Song, Jihyo Kim and KyungTae Lim31933

When Helpfulness Becomes Scientific Sycophancy: Triggering and Mitigating LLM Failures on Flawed Scientific Premises
Yiyuan Zhang, Cailong Hua, Murti Salapaka and Dongyeop Kang 31958

SciTrace: Trajectory-Aware Safety Reasoning for Scientific Discovery Agents
Tanush Swaminathan, Runmin Jiang, Letian Zhang and Min Xu 31988

Beyond Task-Based Evaluation: Toward Situated Benchmarks for Educational LLMs
Xiaolongwang, Tianyi Xu, Rui Zou, Jiawen Zuo, Mingzi Zhang, Keqian Li, Zhe Zhao, Ye Wei and Aimin Zhou 32011

KREL: Automatic Medical Coding via Knowledge-Guided Reasoning over Clinical Evidence with LLMs
Xubin Chen, Yipeng Zhou, Wen Sun, Chengkai Huang, Xiaoming Fu and Quan Z. Sheng . 32024

Mitigating Object Hallucinations in Vision-Language Models via Head-Level Attribution and Counteractive Pruning
Wu Liyan, Xin Luo, Fan Wu, Yingen Liu and Wuxiaobin 32040

Can LVLMs Uncover the Truth Behind Visual Illusions? An Analysis of Perceptual and Reasoning Capabilities
Liangjie Zhao, Jiaqing Lyu, Kexin Tang, Zecheng Fang, Rong Yin, Yulan Hu, Da Li and Jianing Li 32061

Graph Neural Networks for Automatic Term Generation: Defining and Exploring Terminological Graph
Toshihide Fujii, Amir Hazem and Kyo Kageura 32074

Implicit Personality Representations in Humans and LLMs
Yilin Geng, Omri Abend, Eduard Hovy and Lea Frermann 32092

Behavior2Trip: Towards Personalized Travel Planning via User Behavior Trajectory
Zihao Cheng, Yingyu Shan, Hongru Wang, Zeming Liu, Xinyi Wang, Xiangrong Zhu, Yuhang Guo, Wei Lin and Yunhong Wang 32110

LongPIBench: A Long-Context Benchmark for Prompt Injection
Yupei Liu, Yuqi Jia, Neil Zhenqiang Gong and Jinyuan Jia 32132

Agent4Weakness: An Agentic Framework for In-Depth Model Weakness Discovery
Xuanliang Zhang, Chenghao Yang, Zhoufutu Wen, Dingzirui Wang, Ge Zhang, Xiying Zhao, Tianren Feng, Jianpeng Jiao, Jingkai Liu, Zaiyuan Wang, Zuo Wang, Wenya Wu, Zhou Huan, Jin Chen, Wenhao Huang, Qingfu Zhu and Wanxiang Che 32150

MedCalc-Pro: Solving Complex Medical Calculations with LLM Agents
Siran Zhao, Ruihui Hou, Ziyue Huai, Chennuo Zhang and Tong Ruan 32187

Retrieval-Augmented Simplification of Sectioned Biomedical Abstracts
Pascal Mathas, Jan Bakker and Jaap Kamps 32209

FaVOR: LLM-Based Agentic Framework for Factor Mining via Empirical Validation
Hyeonjin Kim, Minseok Kim, Seunghyeon Jung, Sujin Pyo, Huisu Jang and Woojin Lee . . 32220

Decision Rigidity in LLM Disinformation Detection: Behavioral and Representational Evidence
Ming-Jhe Hu and Hung-Yu Kao 32248

Decouple and Orthogonalize: A Data-Free Framework for LoRA Merging
Shenghe Zheng, Hongzhi Wang, Chenyu Huang, Xiaohui Wang, Jiayuan Fan, Shuyue Hu and Peng Ye 32287

ReasonMap: Making Implicit Mathematical Knowledge Explicit in Long Chain-of-Thought Traces
Haodong Wu, Yuhang He and Yongqi Zhang 32308

VizRAG: Enhancing Retrieval-Augmented Generation with Hypergraph Visualization
Yanbin Wei, Yang Chen, Renling Gan, Ziru Liu, Xinyu Fu, Chun Kang, Ning Lu, Rui Liu, Yu Zhang and James Kwok 32327

HuggingBrew: Brewing High-Quality Mid-Training Data from the HuggingFace Community Datasets
Han Zhang, Ningqi Chen, Xiao Liu, Zhenghao Lin, Yeyun Gong, Yan Lu and Hanfang Yang 32346

Transition Overlap: Construction-Time Conditioning for Position-Independent KV Cache Reuse
Liang Xu, Zhichun Wang and Xuetao Tian 32366

PrAM: Offline Checkpoint Selection for Post-Hoc Training Data Attribution
Stephanie Schoch, Sidhya Pathak and Yangfeng Ji 32387

Counterfactual Self-Distillation: A Distribution-Shaping Framework for Entity-Fact Unlearning in Large Language Models
Ying Zhao, Yihui Zhang, Mengyi Deng, Xin Li, Yuan Tian, Wei Wang and Yi Chang 32401

Why Knowing Both Hops Is Not Enough: Understanding Two-Hop Generalization in Language Models
Zili Zhang, Yilin Wang, Heng Wang, Herun Wan and Minnan Luo 32415

Unified Fragment Growing, Linking, and Merging with Energy-Aware Pre-training and Reward-Ranked Alignment
Xuefeng Liu, Songhao Jiang, Tinson Xu, Qinan Huang and Rick L. Stevens 32439

HoloDP: A Multilingual Benchmark Suite and Script-Aware Framework for Document Parsing
Mengxiao Zhu, Wencan Ying, Yingjie Cao, Jie Liu and Jiu Sha 32453

LookBack: Where and How to Score LVLM Responses via Visual Reference Usage
Beomsik Cho, JinHyeong Kim, Dongseok Lee and Jaehyung Kim 32471

LogitsCoder: Towards Efficient Chain-of-Thought Path Search via Logits Preference Decoding for Code Generation
Jizheng Chen, Weiming Zhang, Xinyi Dai, Lingyue Fu, Weiwen Liu, Kounianhua Du, Qingyao Li, Yong Yu and Weinan Zhang 32495

Towards Generalizable Visually Grounded Exploration of Household Devices
Linhao Zheng, Zeming Liu, Wangke Chen, Li Zeng, Wanxiang Che, Heyan Huang and Yuhang Guo 32511

HybridProver: Augmenting Theorem Proving with LLM-Driven Proof Synthesis and Refinement
Jilin Hu, Jianyu Zhang, Yongwang Zhao and Talia Ringer 32542

HyperPathsRAG: Multi-Target Path Approximation over Hypergraph Knowledge Graphs for Multi-Hop QA
Guanhua Zhu, Zirui He, Mengnan Du and Chenxi Yuan 32558

GRAIN: Bridging Name and Narrative Shifts in Real-World Graph Reasoning through Invariance-Rewarded Agentic RL
Zike Yuan, Han Zhang, Jianzhi Yan, Le Liu, Cai Ke, Huozhi Zhou, Jian Xie, Jaron Yin, Yukun Cao, Yue Yu, Hui Wang, Ming Liu and Bing Qin 32584

Counterfactual Selective Imitation for Long-Horizon Agents
Chunlin Yu, Ruiqi Xu, Huan Li and Chuliang Weng 32605

Ada-SQL: Efficient and Adaptive Reasoning for Text-to-SQL
Xiaohan Shen, Xiao Ling and Jie Liu 32619

Beyond Gradient Egalitarianism: Dual-Adaptive C-AUM Framework for Few-Shot Fine-Tuning of Large Language Models
Qiaojuan Peng, Xiong Luo, Yuqi Yuan and Wenbing Zhao 32641

Interpretable Oracle Bone Script Decipherment via Progressive Glyph Analysis with LVLMs
Kaixin Peng, Mengyang Zhao, Haiyang Yu, Teng Fu and Bin Li 32654

NUM2EVENT: Structured Event Inference from Numerical Time-Series
Ninghui Feng, Songning Lai, Xinan Chen, Zheng Chu and Yiyan Qi 32675

CQA-Eval: Designing Reliable Evaluations of Multi-paragraph Clinical QA under Resource Constraints
Federica Bologna, Tiffany Pan, Matthew Wilkens, Yue Guo and Lucy Lu Wang 32694

APEX-Searcher: Hierarchical Credit Assignment with Subgoaling for Agentic Retrieval-Augmented Generation
Kun Chen, Qingchao Kong, Zhao Feifei and Wenji Mao 32708

When AI-Text Detectors Become Evidence: Rethinking Detection for Human Decisions
Jiatao Li, Mao Ye, Xinyu Hu, Cheng Peng, Xunjian Yin and Xiaojun Wan 32727

Personalized Privacy Control in LLMs via Attention Head Intervention
Junseok Kim, Nakyeong Yang and Kyomin Jung 32758

SALUTE: Benchmarking and Adapting LLMs for the Defense Domain
Hyeongcheol Park, Sumin In, Suyeon Myeong, Hogun Park, Sangmin Kim, Moonhyun Lee, Daekyeong Park and Sangpil Kim 32780

ICL4Decomp: Towards Re-executable Reverse Engineering
Xiaohan Wang, Yuxin Hu and Kevin Leach 32819

Agentic Open-Domain Tabular Question Answering with Self-Supervised Multi-Table Retriever
Jianwei Wang, Mengqi Wang, Qing Liu, Xiwei Xu, Zhenchang Xing, Liming Zhu and Wenjie Zhang 32833

Are Fine-Tuned LLM Membership Signals Real? Separating Memorization from Likelihood Shortcuts
Zilong Wang, Yingjie Xu, Yutong Wu, Ruiyang Li, Qinglin Ji, Xueming Tang and Songfeng Lu 32850

DMDIntel: Interpreting Large Language Models via Dynamic Mode Decomposition
Amogh Joshi, Animesh Mukherjee and Sergey Utyuzhnikov 32870

NGDBench: Towards Neural Graph Data Management
LI Yufei, Yisen Gao, Jiaxuan Xiong, Jiaxin Bai, Shijie Zhong, Haoyu Huang, Zhongwei Xie, Hong Ting Tsang and Yangqiu Song 32890

GlobeAudio: A Multilingual Multicultural Benchmark for Naturalistic Evaluation of Large Audio-Language Models
Ryner Tan and Wenxuan Zhang 32924

Asuka-Bench: Benchmarking Code Agents on Underspecified User Intent and Multi-Round Refinement
Xin Wang, Liangtai Sun, Yaoming Zhu, Shuang Zhou, Jiaxing Liu, FengJiao Chen, Lin Qiu, Xuezhi Cao, Xunliang Cai, Licheng Zhang and Zhendong Mao 32938

All Claims Are Equal, but Some Claims Are More Equal Than Others: Importance-Sensitive Factuality Evaluation of LLM Generations
Miriam Wanner, Leif Azzopardi, Paul Thomas, Soham Dan, Benjamin Van Durme and Nick Craswell 32960

"Not in My Backyard": LLMs Uncover Online and Offline Social Biases Against Homelessness
Jonathan A. Karr Jr., Benjamin F. Herbst, Xueyun Li, Matthew L Sisk, Ting Hua, Matthew Hauenstein, Georgina Curto and Nitesh V. Chawla 32982

TIGRIS: Integrating Policy Knowledge Graphs and Theory of Change for Institutionally Grounded Policy Impact Simulation
Haein Yeo, Junghyun Kim, Beejin Son and Kyungsik Han 33008

ATC-Bench: A Simulation-Grounded Benchmark for Evaluating LLM-Generated Air Traffic Control Instructions
Jinjun Cai, Jiaxing Li, Xingyue Wang, Weihuang Zheng, Youyong Kong and Hui Ding . . . 33032

A Tree-based RAG Framework for Evidence-Intensive QA via Adaptive Planning and Topology-Aware Evidence Gathering
Songeun Lee, Kyungjin Min, Injae Na, Lee Suyeong, Chi-young Kim and Woohwan Jung 33056

OpenSkillEval: Automatically Auditing the Open Skill Ecosystem for LLM Agents
Jiahao Ying, Boxian Ai, Wei Tang, Siyuan Liu and Yixin Cao 33083

GraDuLa: Graph-based Dual-branch Language-agnostic Modeling for Misinformation Detection with Cross-lingual Evidences
Atul Kumar Singh and Rahul Mishra 33112

FServe: FP8-Native W4A8KV4 Quantization and Attention Kernel Optimization for Efficient LLM Serving
Daehyeon Baek, Jieun Choi, Jimyoung Son, Jaeyoung Choi, Kyungmin Bin, Seungbeom Choi, Kihyo Moon, Minsung Jang and Hyojung Lee 33132

EviDx: Evidence-Aware Active Diagnosis with Scaffolded LLM Agents
Lihang Zeng, Shaoting Zhang and Xiaofan Zhang 33154

Why LLMs Give In: Conversational Factors and Reasoning Behind Medical Sycophancy
Kaike Ping, Buse Çarık, Caleb Wohn, Xiaohan Ding, Tongshuai Wang and Eugenia Rho . . 33180

SAUL: Sharpness-Aware Augmented-Lagrangian Unlearning
Jaewan Choi, Junyoung Yang and Sangdon Park 33202

Communicating Chess Strategies in Natural Language
Langyuan Cui, Chun Kai Ling and Hwee Tou Ng 33231

Visual Representation Regularization for Multimodal Hallucination Mitigation
Yifan Meng, Enguang Wang, Xusheng Cao, Fei Yang and Xialei Liu 33253

Neuron-Guided Fine-Tuning: Unlocking Efficient Alignment Mechanisms for Large Language Models
Zeyu Wu, Junchao Wu, Shudong Liu, Runzhe Zhan, Xin Chen, Shu Yang, Yichao Du, Longyue Wang, Weihua Luo, Jinsong Su and Derek F. Wong 33268

Plan the Semantics, Then Render the Signal: A Hybrid State-Space Model for Incomplete Multimodal Learning
Ruiting Dai, Zhihao Qin, Yilin Chen, Xianhua Yan, Lisi Mo and Tao He 33292

A Thousand Words or An Image: Studying the Influence of Persona Modality in Multimodal LLMs
Julius Broomfield, Kartik Sharma and Srijan Kumar 33309

Learning Structured Emotional Dynamics in Latent Space via Trajectory Modeling and Controlled Transitions
Agampreet Saini, Bayan Abusalameh, Rahul Kumar Singh and Erik Cambria 33329

Skill-Guided Continuation Distillation for GUI Agents
Zhimin Fan, Hongwei Yu, Yeqing Shen, Haolong Yan, Guozhen Peng, Tianhao Peng, Yudong Zhang, Xiaowen Zhang, Kaijun Tan, Zheng Ge, Xiangyu Zhang and Daxin Jiang 33355

SLMFix: Leveraging Small Language Models for Domain Specific Language Error Fixing with Reinforcement Learning
David Jiahao Fu, Aryan Gupta, Aaron Councilman, Yu-Xiong Wang and Vikram S. Adve . 33374

CabinProAct: Visually Triggered Proactive Intervention for Intelligent Cabin Agents
Chengqing Bu, Mengxue Qu, Mingming Ma, Yuyao Huang, Yao Zhao and Yunchao Wei . 33392

ChipLayoutBench: A Multimodal Benchmark for Vision-Language Models in Semiconductor Layout Understanding
Serim Ryou, Jayeon Yoo, Minjae Jeong, Won Hwa Kim and Seong-Jin Park 33414

SocialReasonBench: A Video-QA Benchmark for Social Reasoning with Counterfactual Narrative Videos
Zheyu Huang, Zijing Shi, Haozhe Luo, Huadong Tang, Mingyu Liu, Meng Fang and Ling Chen 33442

Self- and Other-Labels Induce Bidirectional Bias in LLM Judges
Songeun Chae, Min Kim, Donghoon Jung, Seojin Choi and Seohyon Jung 33466

PatientAct: Theory-Grounded Mental Health Client Simulation
Sahand Sabour, NG Tsz Yam, Yaqian Chen, Guanqun Bi and Minlie Huang 33483

RADAM: A Retrieval-Augmented Detection and Mitigation Framework for Faithfulness Hallucinations
Bin He, Guanghua Li, Yu Liu, Mingyang Zhou, Rui Mao, Naipeng Chao and Hao Liao . . . 33507

SCP-ISA: Spatial Cross-attention Projection for Textual Spatial Semantic Anomaly Detection and Attribution
Bei Xu and Mengqi Ding 33528

Cross-Session Decomposition Attacks: Scaling Risk and Intent-Aligned Retrieval Defense
Disen Liao, Yihan Wang, Freda Shi and Yaoliang Yu 33535

SAFIRE: Safety-Critical Benchmark for Fine-grained Fire and Smoke Understanding in Multimodal LLMs
Pengfei Li, Naufal Suryanto, Sicheng Zhang, Mohammad Alsharid and Muzammal Naseer33565

Relevant Is Not Warranted: Evidence-Force Calibration for Cited RAG
Pin Qian, Su Wang, Yihang Chen, Wenxuan Xu, Xiaoyuan Wang, Qiaolin Yu, Shuhuai Lin, Sipeng Zhang, Junxian You and Xinpeng Wei 33590

LivingRAG: Augmenting Graph RAG with Experience
Yuzhuo Cui, Zongye Zhang and Qingjie Liu 33604

Human Inputs, Model Outputs: Dual Sources of Variability in Stance Mining on Gun Rights/Control
Niranjana Sathish Avilery, Soumyajit Datta, Ashini Anantharaman and Ashiqur R. KhudaBukhsh 33629

SurrogateKV: Representation-Preserving KV Cache Compression for Long-Context LLMs
Junwon Kim, Junghyun Ryu, Farid Talibli, Jungmin So and Youngjae Kim 33643

DyMT-ESB: Dynamic Multi-Turn Evaluation of Social Bias in User-LLM Interactions
Rem Hida, Masahiro Kaneko, Daisuke Oba, Danushka Bollegala and Naoaki Okazaki 33661

Skill-MAS: Evolving Meta-Skill for Automatic Multi-Agent Systems
Hehai Lin, Qi Yang and Chengwei Qin 33685

Not All Components Forget Alike: Component-Level Intervention Profiles in Continual Pretraining
Sangyeol Baek and Suan Lee 33719

Retrieve Few, Verify Many: Segment-Level Evidence Reuse for Long-Form Factuality Assessment
Junjie Liao and Yuxia Wang 33740

Scaling Multi-Agent Systems with Prospect-State Propagation
Zhimei Chen, Mu Chen and Fakhri Karray 33754

Uncovering and Mitigating Aggregation-Induced Reward Hacking in Multi-Reward Reinforcement Learning
Yu Yuan, Yaoyou Fan, Lili Zhao, Guangting Zheng, Kai Zhang, Lu Pan, Ke Zeng and Qi Liu 33767

Black-Box Dataset-Level Source Ranking via Context Mixture Factorization
Milad Fotouhi, Mohammad Taha Bahadori, Payman Arabshahi and Oluwaseyi Feyisetan . 33781

Discerning Minds or Generic Tutors? Evaluating Instructional Guidance Capabilities in Socratic LLMs
Ying Liu, Can Li, Ting Zhang, Mei Wang, Qiannan Zhu, Jian Li and Hua Huang 33802

LIRQuant: An Efficient Locally Isolated Rotation Quantization Method for Large Language Models
Desheng Xu, Hanglin Li, Tianchen Song, Zhiyong Zhao and Kun Zhan 33835

Compressing Long Context into Answer-Aligned Memory Embeddings for LLM Inference
Md Mostafizer Rahman, Md Faizul Ibne Amin, Md. Shahajada Mia, Yutaka Watanobe and Fang Liu 33852

Chart-EARL: Improving Chart Reasoning of Vision LLMs via Early Anchoring in Reinforcement Learning
Gongyao Jiang and Qiong Luo 33872

IALR: Instance-Adaptive Legal Rubric Supervision for Legal Judgment Prediction
Jianbai Zhao, Bibo Cai, Xiao Ding, Xinran Dai, Yu Bao, Bo Xin, Kai Xiong, Bing Qin and Ting Liu 33890

From Base Rollouts to RL Reasoning: A Budgeted Search Perspective
Wenhe Sun, Cunxiang Wang, Zijun Yao and Yixin Cao 33908

Scaled Autoregressive Projection Steering (SAPS): Token-level Steering to Induce Sarcasm and Humblebrag in Text
Sharath Naganna, Ritesh Sur Chowdhury, Dipankar Srirag and Aditya Joshi 33932

Learning to Predict Future-Aligned Research Proposals with Language Models
Heng Wang, Pengcheng Jiang, Jiashuo Sun, Zhiyi Shi, Haofei Yu, Jiawei Han and Heng Ji 33962

Quantifying and Mitigating Default Behaviors in Text-to-Image Models
Royi Rassin, Aviv Slobodkin, Shauli Ravfogel, Yanai Elazar and Yoav Goldberg 33994

SPUR-XMC: Context-Adaptive Spurious Correlation Mitigation for Extreme Multi-Label Classification of Biomedical Literature
Ratri Mukherjee, Shailesh Dahal and Kishlay Jha 34025

HarmMetric Eval: Benchmarking Metrics and Judges for LLM Harmfulness Assessment
Langqi Yang, Yixuan Chen, Kedong Xiu, Tianhang Zheng, Hao Zhou, Wangze Ni, Lei Chen, Zhan Qin and Kui Ren 34046

A Survey on Rubric-guided Reinforcement Learning for Language Models
Zifei Shan and Fangning Shao 34065

Revising Context, Shifting Simulated Stance: Auditing LLM-Based Stance Simulation in Online Discussions
Xinnong Zhang, Wanting Shan, Hanjia Lyu, Zhongyu Wei and Jiebo Luo 34088

MedCPEnv: A Reproducible Cost-Aware Simulated Hospital for Benchmarking and Training Medical LLM Agents
Zixiao Jia, Changjie Hua, Weiqi Zhai, Ruyi Xu, Xue Yang, Han Li, Tianyun Lu, Kailuan Wu, Xiang Zheng, Lin Qu, hu Wei, Bing Zhao and Ruohan Zhang 34108

Post-Training Shifts Confidence: A Three-Stage Analysis of How SFT, RL, and OPD Shape CoT Calibration
Shuhao Li, Guodong DU, Anhao Zhao, Wanyu Lin, Tianyu Yuan and Xiaoyu Shen 34126

CultureVLM: Characterizing and Improving Cultural Understanding of Vision-Language Models
Shudong Liu, Yiqiao Jin, Cheng LI, Qingsong Wen, Lichao Sun, Haipeng Chen, Derek F. Wong and Jindong Wang 34142

Taming the Thinker: Conditional Entropy Shaping for Adaptive LLM Reasoning
Shuyu Wei, Jian Sun, Delai Qiu, Yining Wang, Shengping Liu, Jerry Liang, Ying Fu, Wei Huang and Jitao Sang 34167

Thought-Aware KV Cache Compaction for Reasoning via Adaptive Attention Matching
Yang Liu, Bin Chong, Chongyang Zhang, Hao Zheng, Jiayu Liang and Xu Kefu 34181

Revisiting Task Vectors: A Structural Analysis of Their Limits
Leekyeung Han, Hyeseung JI, Sangwon Jung, Seungwoo Lyu and Paul Hongsuck Seo 34196

Does Reasoning Preserve Alignment? On the Trustworthiness of Large Reasoning Models
Prajakta Kini, Avinash Reddy, Souradip Chakraborty, Satya Sai Srinath Namburi GNVV, Furong Huang, Amrit Singh Bedi and Alvaro Velasquez 34213

From Narrative to Auditable Forecasts: A Structured Scaffold for Agentic Forecasting
Yuanpu Cao, Yongkang Du, Yurui Chang, Lu Lin and Jinghui Chen 34249

RAMP: Regime-Aware Merge-Prune Token Compression for ASR-Oriented Audio-Language Models
Jaehan Jung, Taehan Lee and Hyukjun Lee 34268

RADAR: Accelerate Large Language Model Inference with RL-Based Dynamic Draft Trees
Junjie Ma, Jinlong Li and Jiajun Luo 34289

DeepResearch RewardBench: Evaluating Reward Models for Deep Research Report
Shudong Liu, Yuting Zhong, Guanhua Chen, Xinyi Yang, Junnan Liu, Lidia S. Chao, Jiaheng Liu and Derek F. Wong 34302

Not All Tokens See Equally: Perception-Grounded Policy Optimization for Large Vision-Language Models
Zekai Ye, Xiaocheng Feng, Ruihan Chen, Ziming Li, Haoyu Ren, Kun Chen, Dandan Tu and Bing Qin 34320

Social Caption: Evaluating Social Understanding in Multimodal Models
Leena Mathur, Bhaavanaa Thumu, Gaoussou Youssouf Kebe and Louis-Philippe Morency 34345

ZipRL: Adaptive Multi-Turn Context Compression with Hindsight Response Replay
Zhexin Hu, Li Wang, Xiaohan Wang, Jiajun Chai, Xiaojun Guo, Wei Lin, Xin Zhou and Guojun Yin 34371

ProofFrontier: Turning Verified Proofs into Boundary-hard Tasks for Formal Theorem Proving
Zhengyan Shi, Jianyuan Zhong, Xiaolin Zheng, Liangliang Shi and Qiang Xu 34392

You Frame It: How Conceptual Representations Shape LLM Detection and Reasoning about Antisemitism
Katharina Soemer and Helena Mihaljevic 34405

Learning When to Think: Utility-Guided Adaptive Reasoning for Efficient Text Retrieval
Yingtao Luo, Kangyi Zhao, Shiyi Du and Yubo Li 34424

MMRad-22K: A Structured Multimodal Evidence Dataset for Chest X-ray Report Generation
Yichen Zhao, Zelin Peng, Fenghe Tang, Piao Yang, Yu Huang and Wei Shen 34435

LSH-RL: Long-Short Horizon Reinforcement Learning for Persona-Level Role-Playing Agents
Wenlong Shi, Jianxun Lian, Haiming Qin, Jiwei Zhang, Wei Zhang, Mingyang Zhou and Hao Liao 34453

PrivAgentFlow: Agentic Workflow for Automating Privacy Preservation in Web Agents
Tianyi Tang, Tianyi Ma, Yueming Lyu, Yew-Soon Ong, Ivor Tsang and Haiyan Yin 34471

REALM: Reliable Expertise-Aware Language Model Fine-Tuning from Noisy Annotations
Sajjad Ghiasvand, Mark Beliaev, Mahnoosh Alizadeh and Ramtin Pedarsani 34492

Data-Driven Red-Teaming: Reusing Adversarial Attacks via Metadata-Aware Retrieval
Mateusz Kowalczyk, Marcin Sendera and Sebastian Cygert 34510

STAR-Table: Spatial Textual Anchoring and Reconstruction for Hierarchical Table Reasoning via OCR
Zhoujian Wang, Han Zhang and Hanfang Yang 34529

Training-Free Policy Violation Detection via Activation-Space Whitening in LLM
Oren Rachmil, Avishag Shapira, Roy Betser, Omer Hofman, Itay Gershon, Yuval Elovici, Asaf Shabtai and Roman Vainshtein 34559

ASSAY: Reconciling Stability and Validity in Reference-Free Scientific LLM Judges
Soyeon Hong, Heegyu Kim, Ji Hoon Hong, Jindong Hwang, Juhyeon Ha, Sung Beom Cho and Hyunsouk Cho 34580

Fusian: Multi-LoRA Fusion for Fine-Grained Continuous MBTI Personality Control in Large Language Models
Zehao Chen and Rong Pan 34601

PRISM: Multidimensional Safety Risk Detection and Measurement Beyond Superficial Compliance for LLMs
YIfan Sun, Xiang Zhao, Beining Lin, Feng Huang and You Song 34612

Text Attribution as Continuous Trajectories: Paired AUC Metrics and Mid-Layer Allocation-Effect Decomposition
Zihan Gu, Xinyan Zhang, TianYi Zhang, Han Zhang, Hua Zhang and Yue Hu 34633

From Detection to Refusal: Safer LLMs via Circuit-Guided Weight Scaling
Kuan-Lin Chu, Chung-En Sun and Tsui-Wei Weng 34657

NDD-R1: Reinforcement Learning for Multimodal LLMs to Understand Communication in Non- and Minimally-Speaking Autistic Individuals
Siddhant Bikram Shah and Kristina T. Johnson 34681

Preserving General Capability in Personality-Aligned LLMs via Orthogonal Gradient Projection
Yangfu Zhu, Linyue Wei, Nianwen Ning, Pengpeng Zhou, Chenguang Song, Yuandong Wang and Zhenzhou Shao 34698

EviGen: Predictive Evidence Scaffolding for Verifiable Clinical Rationale Generation
Fengnan Li, Heman Burre, Liwen Sun, Roshni Varma and Matthew M. Engelhard 34715

Mirror, Mirror on the Wall: Prompt Echoing in Instruct Language Models
Inez Okulska, Bartosz Naskręcki, Jan Piotrowski and Tomasz Steifer 34744

Learning to Code by Learning to Verify: Execution-Grounded Co-Evolution of Code Generation and Self-Verification for Reliable Code LLMs
Juyong Jiang, Chansung Park, Jiasi Shen and Sunghun Kim 34751

Cost-efficient Active Learning for Referring Image Segmentation and Grounding
Junbeom Hong, Seonghoon Yu, Hyung Rok Jung, Sundong Kim and Jeany Son 34765

Gaming the Judge: How Unfaithful Chain-of-Thought Can Undermine Agent Evaluation
Muhammad Khalifa, Lajanugen Logeswaran, Jaekyeom Kim, Sungryull Sohn, Yunxiang Zhang, Moontae Lee, Hao Peng, Lu Wang and Honglak Lee 34789

From Memory to Skills: Evidence-Grounded Governance for Long-Horizon LLM Agents
Bo Tang, Yang Zhang, Guomian Zhuang, Wenqiang Wei, Gaoyang Zheng, Lindong Xie, Yanchao Tan, Feiyu Xiong, Qingyu Yang, Edward Chung and Zhiyu li 34826

Is LoRA Robust to Adversarial Attacks? Expectation-Based Regularization for Parameter-Efficient Adversarial Training
Mingze Ni, Yongshun Gong, Xinghao Yang and Wei Liu 34848

HINTBench: Horizon-agent Intrinsic Non-attack Trajectory Benchmark
Jiacheng Wang, Jinchang Hou, Fabian Wang, Ping Jian, Zhonghou Lv and Chenfu Bao . . . 34862

CODEPIVOT: Bootstrapping Multilingual Transpilation in LLMs via Reinforcement Learning with Reduced Reliance on Parallel Corpora
Shangyu LI, Juyong Jiang, Meibo Ren, Sizhe Zhong, Huiri Tan, Yunhao Gou, Xu Han, Chun Yong Chong, Yun Peng and Jiasi Shen 34889

MetaGAI: A Large-Scale and High-Quality Benchmark for Generative AI Model and Data Card Generation
Haoxuan Zhang, Ruochi Li, Yang Zhang, Zhenni Liang, Junhua Ding, Ting Xiao and Haihua Chen 34910

Before the Script, Set the Stage: How Worldview Simulation Amplifies Psychologically Grounded Persuasion in Multi-Turn Jailbreaking
Siyu Chen, Haoran Wang, Xiaojian Li, Yao Huang, Yinpeng Dong and Wei Xu 34932

Frames2LoRA: Parametric Video Internalization for Vision-Language Models
Manan Suri, Sarvesh Baskar and Dinesh Manocha 34951

When Should Queries Be Decomposed? A Stage-Aware Study of Query Decomposition for Multi-Condition Retrieval
Bochao Yin, Xuan Lu, Zhengyu Qi and Xiaoyu Shen 34990

Puzzles to Paradigms: A Survey of Human–Agent Collaboration in Scientific Innovation
Zhenni Liang, Haoxuan Zhang, Haihua Chen, Ting Xiao, Junhua Ding and Yang Zhang . . 35017

TALIM: A Benchmark Dataset for Tajik Academic Language Intelligence Measurement
Azamat Rashidov, Anis Rashidov, Archit Narumanchi and Jong C. Park 35042

Hybrid PIR: Balancing Token Compression and Structural Invariance via Pāṇinian Role Assignment
Kriti Bharadwaj and Gowri Srinivasa 35057

Linearly Readable, Not Robust: How Linear Jailbreak Probes Conflate Benign Register with Adversarial Intent
JaeSeong Kim and Suan Lee 35071

EvoCurr: Self-evolving Curriculum with Behavior Code Generation for Complex Decision-making
Yang Cheng, Weiyu Ma, Zilai Wang, Wenhui Zhu, Tangjie Lv, Yujing Hu, Mohamed Elhoseiny, Yue Deng and Jian Zhao 35089

UniSurgBench: Benchmarking Surgical Video Understanding with Step-wise Reasoning and Workflow Aligned Evaluation
Zhenxin Lei, Mengyang Li, Qian Li, Haozhe Cui, Zhimin Zhao, Hu Han, Shiguang Shan and Xilin Chen 35125

How do World Models and Policies Compose in LLM Agents? A Joint Spectral and Behavioral Account
Ruize Xu, Xiao Yu, Yujin Tang, Chenming Shang and Nikhil Singh 35162

Bridging SLMs and LLMs: A Lightweight Framework for Knowledge Injection
Zihao Li, Xuekong Xu, Yuqing Yu, Lixin Zou, Chenliang Li and Wei Lu 35183

QO-Bench: Diagnosing Query-Operator-Preserving Retrieval over Typed Event Tuples
Mengao Zhang, Xiang Yang, Liu Chang, Tan Tianhui and Ke-Wei Huang 35211

Inside the LLM Word Factory
Benzi Busigin and Yuval Pinter 35230

Evaluating Cross-lingual Knowledge Consistency in Code-Mixed vis-a-vis Indian Languages using IndicKLAR
Debajyoti Mazumder, Divyansh Pathak, Prashant Kodali, Aditya Joshi, Akshay Agarwal and Jasabanta Patro 35250

From Helpfulness to Toxic Proactivity: Diagnosing Behavioral Misalignment in LLM Agents
Xinyue Wang, Yuanhe Zhang, Zhengshuo Gong, Haoran Gao, Fanyu Meng, Zhenhong Zhou, Li Sun, Yang Liu and Sen Su 35278

THRD:A Training-Free Multi-Turn Defense Framework for Jailbreak Attacks on Large Language Models
Zhiqing ma, Zhonghao Xu, Dong Yu, Chen Kang, Changliang Li and Pengyuan Liu 35313

Lost in Localization: Benchmarking Intralingual Cultural Adaptation in Large Language Models
Soumedhik Bharati, Maharaj Brahma and Maunendra Sankar Desarkar 35338

Mining Legal Arguments in U.S. Corporate Case Law
Luis Brena, William Jurayj, Gregory Deyesu, Zaid Al-Huneidi, Andrew Blair-Stanek and Benjamin Van Durme 35359

Patch the Distribution Mismatch: RL-Based Data Rewriting for Stable Downstream SFT
Jiacheng Wang, Zhijie Liu, Ping Jian, Zirong Chen, Ke Ren Liao, Zhen Yang and Zhongbin Guo 35387

Conditional Accuracy Profiles: Diagnosing LLM Judges across Deployment Conditions
Wenqi Li, Bin Liu, Mindi Ruan, Chuanbo Hu, Minglei Yin and Xin Li 35408

Peak Information Flux: A Resource Rational Measure of Language Processing Difficulty
Millan Hamilton and Jiaxuan Li 35428

KSAFE-MM: A Multimodal Safety Benchmark via Localized Contextualization for Korean Cultural Risks
Yongwoo Kim, Sojung An, Yunjin Park, Jungwon Yoon, Dujin Lee, HyunBeom Cho, Jaewon Lee, Wonhyuk Lee, Youngchol Kim, JeongYeop Kim and Donghyun Kim 35442

Evaluating Harmful Memorization in LLM Code Generation
Lizhe Zhang, Wentao Chen, Li Zhong, Letian Peng, Zilong Wang and Jingbo Shang 35469

DisasterBench: Benchmarking LLM Planning under Typed Tool Interface Constraints
Zhitong Chen, Kai Yin, Weifeng Zhang, Zhiyuan Wang, Xiangjue Dong, Chengkai Liu, Zhewei Liu, Yiming Xiao, Ali Mostafavi and James Caverlee 35492

ExpRL: Unified Non-Parametric and Parametric Experience Learning for LLM Agent Evolution
Zhuoran Jin, Lingshuai Wang, Yupu Hao, Tianyi Men, Zhitao He, Daojian Zeng, Kang Liu and Jun Zhao 35524

KITE: Kernelized and Information Theoretic Exemplars for In-Context Learning
Vaibhav Singh, Soumya Suvra Ghosal, Kapu Nirmal Joshua, Soumyabrata Pal and Sayak Ray Chowdhury 35547

Grove MoE: Towards Efficient and Superior MoE LLMs with Adjugate Experts
Haoyuan Wu, Haoxing Chen, Xiaodong Chen, Zhanchao Zhou, Tieyuan Chen, Yihong Zhuang, Guoshan Lu, Zenan Huang, Junbo Zhao, Lin Liu, Zhenzhong Lan, Bei Yu and Jianguo Li 35569

Shopping Companion: Benchmarking and Training LLM Agents for Long-Horizon Preference-Grounded E-Commerce Tasks
Zijian Yu, Kejun Xiao, Huaipeng Zhao, Tao Luo and Xiaoyi Zeng 35582

MPIB: A Benchmark for Medical Prompt Injection Attacks and Clinical Safety in LLMs
Junhyeok Lee, Han Jang and Kyu Sung Choi 35610

Symbolic Mechanistic Data Attribution: Tracing Training Influence to Learned Behavioral Policies
Reza Habibi, Darian Lee and Magy Seif El-Nasr 35629

CC-Mediation: Evaluating Large Language Models for Cross-Cultural Conflict Mediation
Suhyun Lee, Wenxuan Zhang, W. Quin Yow and Yang Deng 35662

Video-Language Entailment for Video Similarity Estimation
Benedetta Liberatori, Lorenzo Vaquero, Elisa Ricci and Paolo Rota 35690

PACE-Bench: Benchmarking Physics Adaptation via Code Evolution in Dynamic Environments
Yuhao Zhan, Bingxiang He, Zecong Tang and Chaojun Xiao 35707

SubT: Subspace Tuning for Few-shot Generalization of Audio-Language Models
Jaehyuk Jang, Kangwook Ko, Wonjun Lee and Changick Kim 35754

VIBE-Bench: Evaluating Personalized Large Language Models When Profiles Don't Mean Preferences
Yiwen Jiang, Yang Deng, Stephanie Fong, Zimu Wang, Yaling Shen, Wei Feng, Hongxi Yang, Xiangyu Zhao, Zhongxing Xu, Deval Mehta, Xuelian Cheng and Zongyuan Ge 35771

Calibration-Free Early Stopping for Chain-of-Thought Reasoning via E-processes
Dong-Jae Lee, Gyojin Han and Junmo Kim 35798

Mem-ABC: Attitude-Aware Agent Memory for Personalized Question Answering
Yinghan Shi, Kai Fu, Quanyou Chu, Xinyu Liu, Ming Du, Hongya Wang, Xiaojun Meng, Jiansheng Wei, Yanghua Xiao and Bo Xu 35815

Beyond Static Dialogues: Benchmarking Realistic, Heterogeneous, and Evolving Long-Horizon Memory
Han Zhang, Zihao Tang, Xin Yu, Xiao Liu, Yeyun Gong, Haizhen Huang, Yan Lu, Weiwei Deng, Feng Sun, Qi Zhang and Hanfang Yang 35836

Recursive Reasoning or Statistical Extrapolation? In-Context Learning in Multi-Agent Interdependent Decision-Making
Yu Liu, Wenwen Li, Yifan Dou and Guangnan Ye 35873

MC-CXR: A Multi-Context Chest X-ray Benchmark for Context-Induced Disruption in Vision-Language Models
Junhyeok Lee, Songsoo Kim and Kyu Sung Choi 35896

OrcaJack: Hijacking Agent Orchestration in Open Multi-Agent Systems
Xiang Li, Haining Yu, Wenbo Pan, Hongli Zhang and Xiaohua Jia 35911

Affordance20Q: Evaluating Affordance Reasoning from Physical Properties
Yifan Jiang, Meige Yang, Zitong Li and Jay Pujara 35930

Beyond My Own Shadow: Collaborative Latent Prompting for Personalized Question Answering
Yinghan Shi, Kai Fu, Xinyu Liu, Quanyou Chu, Ming Du, Hongya Wang, Yanghua Xiao and Bo Xu 35949

TAP-RAG: Task-Aware Policy Control for Long-Document Multimodal Question Answering
Zhong Ji, Keqi Jin, Yan Zhang and Jiasheng Li 35966

AudioProcessBench: Benchmark for Identifying Process Errors in Audio-Grounded Reasoning
Xiangyu Zhao, Junyu Yan, Yaling Shen, Stephanie Fong, Zimu Wang, Yiwen Jiang, Qingyang Xu, Jiahe Liu, Dominic Dwyer and Zongyuan Ge 35981

SAGE: Sink-Aware Guided Emphasis for Visual Grounding in Vision-Language Decoders
Jeonghyo Song and YoungJoon Yoo 36004

AutoHealth: Uncertainty-Aware Multi-Agent Framework for Autonomous Health Prediction
Tong Xia, Weibin Li, Yu Liu, Liu Gang and Yong Li 36027

PTCG-Bench: Can LLM Agents Master Pokémon Trading Card Game?
Dongdong Hua, Yifei Sun, Renhong Huang, Chunping Wang, Feng Gao and Yang Yang . . 36084

All Leaks Count, Some Count More: Interpretable Temporal Contamination Detection and Mitigation in LLM Backtesting
Zeyu Zhang and Bradly Stadie 36100

MVC-Bench: Benchmarking Calibration of Medical Vision-Language Models
Ashshak Sharifdeen, Shihab Aaqil Ahamed, Ufaq Khan, Muhammad Akhtar Munir, Sujair Ibrahim, Mohamed Rafeek Mareer Ahamed, Yutong Xie, Imran Razzak and Muhammad Haris Khan36128

Dummy Backdoor as a Defense: Removing Unknown Backdoors via Shared Internal Mechanisms for Generative LLMs
Kazuki Iwahana, Masaru Matsubayashi, Takuma Koyama, Toshiki Shibahara, Kenichiro Ominato and Akira Ito 36163

Phonological Abstraction and Exception Memorization in Noisy Tone Learning
Yi Yang, Yiming Wang and Jiahong Yuan 36181

Agent4ExpRec: Towards Generating and Evaluating User-Centric Recommendation Explanation
Yayu Hou, Dongjin Yu, Dongjing Wang and Changgeng Li 36200

Immersion in the GitHub Universe: Scaling Coding Agents to Mastery
Zhao Jiale, Guoxin Chen, Fanzhe Meng, Minghao Li, Jie Chen, Hui Xu, Yongshuai Sun, Xin Zhao, Ruihua Song, Yuan Zhang, Peng Wang, Cheng Chen, Ji-Rong Wen and Kai Jia 36216

The Geometry of LLM-as-Judge: Why Inter-LLM Consensus Is Not Human Alignment
Sourabrata Mukherjee, Hamna Abid, Kalika Bali and Sunayana Sitaram 36237

G-STAR: End-to-End Global Speaker-Tracking Attributed Recognition
Jing Peng, Ziyi Chen, Haoyu Li, Yucheng Wang, Mengtian Li, Duo MA, Yu Xi, Yunfan Du, Kai Yu and Shuai Wang 36267

Is Reasoning Always Useful? Rethinking Reasoning Utility in Universal Multimodal Embeddings
Wenxiao Fan, Jingling Fu, Luohang Liu, Xinyuan Shan, Lichen Ma, Yu He, Junshi Huang, Yan Li and Kan Li 36277

ToolSmith: An Automatic Tool-Generation Framework for Multimodal Search Agents
Rui Sun, Junhua Huang, Luca Zancato, Md Mofijul Islam and Kai-Wei Chang 36305

ThinkFuse: Trajectory-Aware Test-Time Fusion for Small Reasoning Models
Myunghoon Kang, Jungseob Lee, Jaehyung Seo and Heuiseok Lim 36325

Adversarial Robustness of Retrieval-Based Fact-Checking: Detecting and Mitigating Evidence Contamination
Alamgir Munir Qazi, John Philip McCrae and Jamal Abdul Nasir 36342

Unknown Unknowns: Do Hidden Intentions in LLMs Evade Detection?
Devansh Srivastav, David Pape and Lea Schönherr 36356

ThinkRetrieve: Retrieval-Augmented Reasoning Traces for Test-Time Scaling
Vaibhav Singh, Soumya Suvra Ghosal, Sarvesh Gharat, Soumyabrata Pal, Ramasuri Narayanam and Dinesh Manocha 36409

ASTER: Agentic Scaling with Tool-integrated Extended Reasoning
Xuqin Zhang, Quan He, Zhenrui Zheng, Zongzhang Zhang, Xu He and Dong Li 36427

PEARL: Perturbation Efficient Alignment Group Relative Reinforcement Learning
Yue Huang, ChenZhuo Zhao, Zhi Yang, Pu Zhao, Fangkai Yang, Lu Wang, Mengqiao Liu, Qingwei Lin, Saravan Rajmohan and Dongmei Zhang 36452

Assessment of Generative Named Entity Recognition in the Era of Large Language Models
Qi Zhan, Yile Wang and Hui Huang 36475

Controllable Affective Generation via Latent Vector Steering
Xixian Yong, Siyuan Chang, Yingying Zhang, Xian Wu and Xiao Zhou 36494

I Wrote This Because of That: Interpretable Radiology Report Generation via Visual Prototypes
Marco Salmè, Federico Siciliano, Fabrizio Silvestri, Paolo Soda, Rosa Sicilia and Valerio Guarrasi 36511

Co-ReAct: Rubrics as Step-Level Collaborators for ReAct Agents
Jiazheng Kang, Bowen Zhang, Zixin Song, Jiangwang Chen, Yuan Shen, Yang Liu and Guanjun Jiang 36528

QatarGuard: A Culturally-Aware Safety Benchmark for Arabic and Multilingual LLMs
Wajdi Zaghouani, Kholoud Khalil Aldous, Md. Rafiul Biswas and Mohammed Abdelkreem36539

GRZO: Group-Relative Zeroth-Order Optimization for Large Language Model Fine-Tuning
Liyan Tan, Yequan Zhao, Yifan Yang, Ruijie Zhang, Xinling Yu and Zheng Zhang 36551

From Enumeration to Principle: Robust Prompt Optimization via Semantic Reflection and Multi-Objective Evolution
Gang Xie, Dongming Shan, Kai Li and Ziyuan Huang 36582

Retrieval Myopia: Why RL Search Agents Fail to Read What They Find
Hyeondong Woo, Jaeho Han, Trung X. Pham and Junyeong Kim 36608

Omanic: Towards Step-wise Evaluation of Multi-hop Reasoning in Large Language Models
Xiaojie Gu, Sherry T. Tong, Aosong Feng, Sophia Simeng Han, Jinghui Lu, Yingjian Chen, Yusuke Iwasawa, Yutaka Matsuo, Chanjun Park, Rex Ying and Irene Li 36623

Confidence-Gated Transductive Test Generation for Code Reranking
Sungjae Lee, Youngsik Yoon, Seockbean Song, Siwei Wang, Wei Chen and Jungseul Ok . 36648

Collaborative OS of Multi-Agent Systems
Jiazheng Kang, Ruochen Zhao, Dongying Lin, Gezi Li and Ting Bai 36661

QUIETT: Query-Independent Table Transformation for Robust Reasoning
Gaurav Najpande, Tampu Ravi Kumar, Manan Roy Choudhury, Neha Valeti, Yanjie Fu and Vivek Gupta 36675

Thinking at the Right Size: Amortized Distillation Across Post-Trained LLMs
Yan Zhou, Sara Kangaslahti, Jonathan Geuter, Nihal V. Nayak, Marco Fumero, Francesco Locatello and David Alvarez-Melis 36697

Code-Switching Reveals Language Anchoring in Multilingual LLMs
Jeonghyun Park, Seunghyun Yoon, Yonghyun Jun and Hwanhee Lee 36729

Back to the Future: Forward-Only Committor Projection for Trajectory-Level Collapse Attribution in RLVR Reasoning
Jaeeun Jang, Hansle Lee and Sangmin Kim 36769

CHILLGuard: Towards Fine-Grained Chinese LLM Safety Guardrail with Scalable Data Construction and Model-aware Preference Alignment
Wenbo Yu, Bohua Wang, Hao Fang, Kuofeng Gao, Jingru Zeng, Xiaochen Yang, Tianyi Zhang, Xiaoxiao Ma, Jiawei Kong, Hao Wu, Bin Chen, Shu-Tao Xia and Min Zhang 36802

EGTR-Review: Efficient Evidence-Grounded Scientific Peer Review Generation via Multi-Agent Teacher Distillation
Xinpeng Qiu, Wang Yihu, Zhifeng Liu, Xiaochen Wang and Jimin Wang 36828

DrDEE: A Deconstruction-Reconstruction Framework for Document-Level Event Extraction
Kuofei Fang, Yuting Wei, Zexuan Wan, Bo Wang, Siyan Wu and Bin Wu 36857

Online Language Adaptive Sampling for Better Distributed Cross-lingual Gains
Quang Phuoc Nguyen, Félix Gaschi, David Anugraha, Santiago Martínez Novoa and En-Shiun Annie Lee 36873

No Action Without a NOD: A Heterogeneous Multi-Agent Architecture for Reliable Service Agents
Zixu Yang, Hang Zheng, Nan Jiang, Zhiyang Tang, Situo Zhang, Xiaobao Wu, Lu Chen and Kai Yu 36892

ArabicDialectSafety: A Dialect-Aware Benchmark for Arabic Content Safety Classification
Wajdi Zaghouani, Md. Rafiul Biswas, Kholoud Khalil Aldous and Mabrouka Bessghaier . 36924

Internalize the Temperature: On-Policy Self-Distillation as Policy Reheater for Reinforcement Learning
Xuewei Yang, Jiachen Yu, Jie Wu, Shaoning Sun, Junjie Wang and Yujiu Yang 36939

MEETINGQA2: A Generative Speaker-Centric Benchmark for Multi-Party Dialogue Question Answering
Yuhan Chen, Yanling Li, Bowei Zou, Yifan Fan and Yu Hong 36957

GuardianAgent: Policy-Conditioned Risk-Adaptive Anonymization with Verified Adversarial Escalation
Ruiyi Yang, Gayathri Lihinikaduarachchi, Rahat Masood, Flora D. Salim and Salil S. Kanhere 36968

SafeBranch: Branch-Pair Safety Alignment for Embodied Agents
Hyunse Lee, Jiwoo Jung, Haneul Lee, Kyochul Jang, Youngjae Yu and Woojin Lee 36985

A Geometry-Aware Depth Function for Robust Unsupervised Textual Anomaly Detection
Pierre Fihey, Matthieu Labeau and Pavlo Mozharovskyi 37011

Two-Stage Reinforcement Learning for Sound and Adversarial Test Generation in Code LLMs
Jiacheng Xu, Wentao Zhang, Zhiyi Lyu, Fuxiang Zhang, Chaojie Wang, Yang Liu and Bo An 37031

What do Reward Models Memorize?
Ivo Verhoeven, Pushkar Mishra and Ekaterina Shutova 37052

DyCo: Dynamic Cognitive Role Coordination for Open-Ended Multi-Agent Research Synthesis and Verification
Zhiyu Zhang, Leheng Wu, Yijun Mo, Mingchu Zhong, Yulin Zhang and Chen XI 37087

Measuring Human-Like Bias in LLMs? A Critique of Human-Derived Bias Constructs in LLM Evaluation
Antonela Tommasel and Markus Schedl 37111

RELOOP: Recursive Retrieval with Multi-Hop Reasoner and Planners for Heterogeneous QA
Ruiyi Yang, Hao Xue, Imran Razzak, Hakim Hacid and Flora D. Salim 37120

GREAT: Generalizable Backdoor Attacks in RLHF via Emotion-Aware Trigger Synthesis
Subrat Kishore Dutta, Yuelin Xu, Piyush Pant and Xiao Zhang 37139

Compliance, Capability, and Conflict: Benchmarking Multimodal LLMs under System Messages
Juan Yeo and Geewook Kim 37171

Trajectory Self-Feedback in Diffusion Language Models: A Systematic Survey
Yash Shah, Abhijit Chakraborty and Vivek Gupta 37189

Navigating Intention Reasoning Trees to Guide Agentic Preference Optimization
Ziben Yue, Hongyu Yang and Weixin Li 37209

SVote-2S: Input-Side Evidence Voting for Two-Stage Self-Consistency in Multi-Hop QA
Guo Ouyang, Ge Wang, Tao Zhu, Yi Zou, Xuehao Yin, Guoliang Li and Fuchuan Ni 37229

Crowd Values: How Human Values Reshape LLM-Agent Communities
Xiangxu Zhang, Jiamin Wang, Qinlin Zhao, Hanze Guo, Linzhuo Li, Jing Yao, Xiaoyuan Yi, Xing Xie and Xiao Zhou 37250

PaperScout: An Autonomous Agent for Academic Paper Search with Process-Aware Sequence-Level Policy Optimization
Tingyue Pan, Jie Ouyang, Mingyue Cheng, Qingchuan Li, Zirui Liu, Daoyu Wang, Mingfan Pan, Shuo Yu, Qi Liu and Enhong Chen 37294

Leaving the Past Behind: Breaking Personalization Inertia in Long-Term LLM Agents
Wenhao Yuan, Chenchen Lin, Jian Chen, Weijie Liu, Jinfeng Xu, Shuo Yang and Edith Cheuk-Han Ngai 37311

RLPgR: Reinforcement Learning from Dependency-Aware Progress Rewards
Shaofan Liu, Tingyi Yang, Shihan Dou, Zhiheng Xi, Junjie Ye, Junlin Shang, Yuming Yang, Shaowen Wang, Jiayi Chen, Jiazheng Zhang, Guoqiang Zhang, Huaibing Xie, Pluto Zhou, Tao Gui, Qi Zhang and Xuanjing Huang 37334

Enhancing Emotional Support Conversation with Plug-in Modules for Sequential Strategy Modeling and Strategy-Specialized Generation
Yerin Park, Junhyeok Park and Junsu Cho 37365

Multimodal Boolean Logic Culture Retrieval
Jian Zhang, Shuheng Hu, Jiutian Chang, Junyi Guo, Huanda Lu, Qiufeng Wang and Fangyu Wu 37380

Neither Here Nor There: Cross-Lingual Representation Dynamics of Code-Mixed Text — A Hindi–English Case Study
Debajyoti Mazumder, Divyansh Pathak, Prashant Kodali and Jasabanta Patro 37390

CL-OCR: Fusing Layout Analysis and Adapting Recognition for Document Parsing in the Wild
Haijie Yuan, Yiming Zhang, Dekang Ran, Yi Chen, Xinming Wang, Jun Cao, Wenzhuo Liu, Xiao-Hui Li, Fei Zhu, Fei Yin and Cheng-Lin Liu 37416

Unveiling Latent Cognitive Bias Constructs in Large Language Models
Bezawit Lake Tilaye, Huw Day, Chris McWilliams and Telmo M Silva Filho 37438

IHDec: Divergence-Steered Contrastive Decoding for Securing Multi-Turn Instruction Hierarchies
Nicole Geumheon Liu, Haeun Jang, Yonghyun Jun and Hwanhee Lee 37459

Stay Within Your Bounds: Distance-Guided Decoding for Guaranteed Context-Free Grammar Compliance
Vincenzo Collura, Karim Tit, Eleonora Giunchiglia, Mike Papadakis and Maxime Cordy . 37479

Remember and Reweight: Enhancing Multi-Agent Debate with Experience Memory and Confidence Estimation
Xuanfa Jin, Zhijian Ma, Yongcheng Zeng, Xinyu Cui, Haifeng Zhang and Jun Wang 37493

Retrieval Evaluation Under Temporal Drift in Technical Documentation
Nathan Kuissi, Suraj Subrahmanyan, Nandan Thakur and Jimmy Lin 37517

ATLAS: Exploring Scalable Activation-Space Data Attribution
Linda Cai, Xinyan Hu, Alexander Pan, Socrates Osorio, Pratyay Pandey, Michael I. Jordan and Jacob Steinhardt 37528

PLANE: Plug-and-Play Fine-Grained Visual Experts for Hallucination Mitigation in Large Vision-Language Models
Jianghao Yin, Lang Yu, Qianxi Yu, Qin Chen, Jie Zhou, Zhikai Lei and Liang He 37559

R^3 -MDT: A Retain-Rerank-Revise Framework Based on Uncertainty and Consistency for Medical Decision Tree Extraction
Boyu Qiu, Xiaolong Weng, Zehua Wang, Buzhou Tang, Ying Xiong and Min Zhang 37577

Ancient-Bench: A Comprehensive Multi-millennial, Multi-medium, and Multi-script Benchmark for Ancient Chinese Artifact Text Recognition
Hiuyi Cheng, Nuo Xu, Yuyi Zhang, Xuhan Zheng, Wei Pan, Jing Zhang, Dezhi Peng, Minghui Liao, Teng Yihua, Jihao Wu, Haoyu Ren and Lianwen Jin 37595

A Circuit for Plural Reference: How LLMs Represent and Retrieve Singular and Plural Entities
Dang Thi Thao Anh, Rick Nouwen and Massimo Poesio 37627

SpikeXL: Language-Agnostic Spike Representations for Few-Shot Cross-Lingual Speech Command Recognition
Vu Son Tung, Minh-Thuan Bui and Duc-Tho Mai 37645

Learning to Search: A Decision-Based Agent for Knowledge-Based Visual Question Answering
Zhuohong Chen, ZhengXian Wu, Yunyao Yu, Hangrui Xu, Zirui Liao, Zhifang Liu, Xiangwen Deng, Peng Jiao and Haoqian Wang 37658

PRFC: Learning Progressive Recovery from Function Calling Errors for Tool-Using LLMs
Beibei Yu, Tao Shen and Ling Chen 37685

EDWARD-NMT: Elastic Dimensions With Adaptive Routing and Dual-consistency for Neural Machine Translation
Edward Nguyen, Trung Thai Tran, Kiet Anh Hoang, Duy Hoang Vien, Duy Vo, Khanh Le and Vu Le Tuan 37707

Predictive Language Grounding for Event-Driven Time Series Forecasting via Dual-Path Event Injection
Jintao Wu, Xin Xue, Yabin Cao, Peng QI, Su Zhao, Shuai Chen, Wei Lin, Tianyu Chen and Haoyi Zhou 37728

HalluPeer: A Taxonomy-driven Benchmark for Detecting Hallucinations in Scientific Peer Reviews
Tzu-Ling Lin, Dong-Ting Yao, Teng-Fang Hsiao, Wei-Chih Chen and Hong-Han Shuai . . . 37746

Learning from Own Solutions: Self-Conditioned Credit Assignment for Reinforcement Learning with Verifiable Rewards
Yingyu Shan, Yuhang Guo, Zihao Cheng, Zeming Liu, Xiangrong Zhu, Xinyi Wang, Jiashu Yao, Wei Lin, Hongru Wang and Heyan Huang 37784

Simple Beats Spend: Cost-Performance Trade-Offs for Multi-Agent Machine Translation of Financial Long Documents
Arturo Oncevay, Zhijin Guo, Elena Kochkina, Damiano Fassina, Simerjot Kaur, Charese Smiley, Xiaomo Liu and Manuela Veloso 37804

OntoDiagnose: Ontology-Grounded Reasoning for Medical Question Answering with Language Models
Pratik Jalan, Mukul Joshi and Kshitij Sharad Jadhav 37819

Long Listening Thoughts: Eliciting Open Auditory Reasoning with Deliberative Perception and Cognitive Refinement
Jaeyeon Kim, Chao-Han Huck Yang, Luoyi Zhang, Chan-Jan Hsu, Fernando Ruiloba Portilla, Jinchuan Tian, Shinji Watanabe and Carlos Busso 37844

Evidence-Guided Bidirectional Iterative Reasoning for Interpretable Personality Recognition
Tinghao Chen, Yuxuan Song, Mengyu Xiang, Qiudan Li, Shu Wu and Daniel Dajun Zeng 37877

UltraVoice: Scaling Fine-Grained Style-Controlled Speech Conversations for Spoken Dialogue Models
Wenming Tu, Guanrou Yang, Ruiqi Yan, Wenxi Chen, Ziyang Ma, Yipeng Kang, Kai Yu, Xie Chen and Zilong Zheng 37897

EvolSQL: Structure-Aware Evolution for Scalable Text-to-SQL Data Synthesis
Xuanguang Pan, Chongyang Tao, Jiayuan Bai, Jianling Gao, Kai Hua, Ge Zhang and Shuai Ma 37915

Think Wider: Mitigating Latent Rank Collapse in Implicit Chain-of-Thought Reasoning
Yuwen Hao and Menglin Yang 37936

Adaptive Test-Time Compute Allocation for Block Diffusion Language Models in Complex Reasoning
Yi Lu, Deyang Kong, Jianing Wang, Linsen Guo, Xue Wang, Qi Guo, Tao Gui, Xuanjing Huang, Wei Ye, Shikun Zhang and Wei Wang 37952

TrajToM: Evaluating Theory of Mind over Mental State Trajectories in Social Dialogues
Chenglong Wu, Xi Wang, Jinzhi Liao and Xiang Zhao 37968

Beyond Token-Level Guidance: Inference-Time Alignment of Specialized LLMs via Cross-Family Representation Steering
Jin Gan, Xin Li and Jun Luo 37991

DualES: Internalizing Query-Text Duality for Text-to-ES via Bidirectional Reasoning-Enhanced Supervision
Dongge Xue, Yuxiang Wang and Tong Ruan 38006

IWP: Token Pruning as Implicit Weight Pruning in Large Vision Language Models
Dong-Jae Lee, Sunghyun Baek and Junmo Kim 38025

SVD-Attention: Efficiently Analyzing Attention Blocks in Transformers
Debopam Sanyal, Hongjie Chen, Riya Elizabeth Kuriyan, Arth Tak, Divya Kothandaraman, Akshay Mehra, Joshua Kimball and Alexey Tumanov 38048

Variance-Adaptive Muon: Pre-Orthogonalization Variance Modulation for Efficient Language Model Pretraining
Jingru li, Yibo Fan and Huan Li 38060

StoryLens: Preference-Aligned Story Rewriting via Context-Aware Narrative Enrichment
Hanwen Cui, Yuting Mei, Yuhang Fu, Dingyi Yang and Qin Jin 38082

Spatial Reasoning in Large Vision-Language Models via Egocentric-to-Allocentric Transformation
Yoonji Kim, Jieun Kim and Sung-Bae Cho 38100

Retrieval-Augmented Agentic Rubric Generation for Reliable Medical Response Evaluation
Yinzhu Chen, Abdine Maiga, Hossein A. Rahmani and Emine Yilmaz 38119

Make an Offer They Can't Refuse: Grounding Bayesian Persuasion in Real-World Dialogues without Pre-Commitment
Buwei He, Yang Liu, Zhaowei Zhang, Zixia Jia, Yang Yu, Huijia Wu, Zhaofeng He, Zilong Zheng and Yipeng Kang 38143

MedFeat: Model-Aware and Explainability-Driven Feature Engineering with LLMs for Tabular Prediction
Zizheng Zhang, Yiming Li, Justin Xu, Jinyu Wang, Rui Wang, Lei Song, Jiang Bian, David W Eyre and Jingjing Fu 38173

Dual-Bias Load Balancing: Mitigating Expert-Bias Staleness in Auxiliary-Loss-Free Mixture-of-Experts
Yuxin Tian, Hailin Zhao, Jia Liu and Zhiqiang Zhang 38193

Beyond Objective Equivalence: Constraint Injection for LLM-Based Optimization Modeling on Vehicle Routing Problems
Xizi Luo, Changhong He, Dongdong Geng, Chenggong Shi and Yu Mei 38207

Beyond Scalar Scores: Exploring LLM-based Metrics for Clinical Significance Evaluation in Radiology Reports
Qingyu Lu, Ruochen Li, Liang Ding, Yufei Xia, Youxiang Zhu and Dacheng Tao 38237

Synthetic Data Characterization via Training Dynamics
Irene Lago, Ana Ezquerro and David Vilares 38265

STARS: State-Aware Reasoning Stopping via Hidden-State Spectral Analysis
Aaron T. Cheng, Junda Wu, Sheldon Yu, Rui Wang, Ryan A. Rossi, Sizhe Zhou, Jiawei Han, Julian McAuley, Lina Yao and Tong Yu 38289

Do No Harm: Exposing Hidden Vulnerabilities of LLMs via Persona-based Client Simulation Attack in Psychological Counseling
Qingyang Xu, Yaling Shen, Stephanie Fong, Zimu Wang, Yiwen Jiang, Xiangyu Zhao, Jiahe Liu, Zhongxing Xu, Vincent Lee and Zongyuan Ge 38307

FEPO: Future-Entropy Policy Optimization with Selective Update Modulation
Cunjie He, Tianhao Cheng, Zhiqi Li, Jiaxin Gu, Xiaoshuang Huang, Jiayin Cai, Xiaolong Jiang, Yao Hu, Jing Xu, Yuan Qi and Zenglin Xu 38326

ConsistKV: Layer Adaptive KV Cache Compression via Cross Layer Consensus
Dongjie Xu, Kai Qian, Weijie Shi, Fenglei Jin and Jiajie Xu 38344

Text-Conditioned Static-Dynamic Composition for Composed Video Retrieval
Yulin Zhang, Haitian Zheng, Zhengwei Yang, Hao Li and Zheng Wang 38356

Patterns of Priming in Production: Lexical, Semantic and Structural Alignment in Language Model Generation
Giulia Pucci, Ruizhe Li and Arabella Sinclair 38374

MppReasoner: Reasoning-Enhanced Large Language Models for Molecular Property Prediction
Jiaxi Zhuang, Yaorui Shi, Jue Hou, Mingwei Ye, Mingjun Xu, Yuming Su, Linfeng Zhang, Ying Qian, Linfeng Zhang, Guolin Ke and Hengxing Cai 38399

From Discovery to Internalization: Length-Adaptive Policy Optimization for Efficient Reasoning
Xingyu Wu, Yuchen Yan, Shangke Lyu, Linjuan Wu, Yiwen Qiu, Yongliang Shen, Weiming Lu, Jian Shao, Jun Xiao and Yueting Zhuang 38422

When Helpfulness, Harmlessness, and Honesty Remember They Are on the Same Team
Gautam Siddharth Kashyap, Mark Dras and Usman Naseem 38443

MCPlanner: Hierarchical Adaptive Planning-Based Medical Consultation with Large Language Models
Qing Feng, Zhihao Jia and Junwen Duan 38455

On the Universality of the Inner Product as a PMI Estimator in Embedding Models
Enrique Amigo, Fernando Giner and Julio Gonzalo 38472

OmniPhys: A Unified Multimodal Benchmark for Physics Understanding and Generation from Chinese Educational Corpora
Hao Chen, Yumin Lin, Nadila Yushanjiang, Xin Lin and Min Zhang 38481

KoNeoBench: A Curated Evaluation Dataset for LLM Understanding of Korean Neologisms
Soha Lee, Soojin Lee, Heesung Yang, Hyunju Song, Hyunji Lee, Jinsan An, Jeongwan Shin, Jin Hyun Park, Jun Lee, Hyeyoung Park and Kilim Nam 38499

TriLens: Per-Layer Logit-Lens Entropy for White-Box Hallucination Detection
Bohan Yang, Yijun Gong, Zhi Zhang, Ge Zhang, Wenpeng Xing and Meng Han 38536

MemoryCD: Benchmarking Long-Context User Memory of LLM Agents for Lifelong Cross-Domain Personalization
Weizhi Zhang, Xiaokai Wei, Wei-Chieh Huang, Zheng Hui, Chen Wang, Michelle Gong and Philip S. Yu 38554

Beyond ID Embeddings: Process-Grounded Language Modeling for Cognitive Diagnosis
Minghang Liu, Yuanzhuo Wang, Qiang Qiu, Huawei Shen and Xueqi Cheng 38576

ShardMemo: Scope-Before-Routing for Agentic Memory Retrieval
Yang Zhao, Chengxiao Dai, Mengying Kou, Yue Xiu and Dusit Niyato 38596

SoVTR: Set-of-Vision-Text Rewards for Attribute-Grounded Emotion Recognition in VLLMs
Qixuan Zhang, Zhifeng Wang, Donghao Zhang, Jiyang Zheng, Wenjia Niu, Ramesh Sankaranarayana, Sabrina Caldwell, Tom Gedeon, Yang Liu and Zhenyue Qin 38611

Ethical Imprints in LLMs: From Corpus Ethics to ICL, SFT, and Their Interaction on Deception Dilemmas
Jiafeng Wang, Jiale Chen, Zheng Luo, Xinyi Yan and Lingpeng Kong 38630

Hallucination-Aware Multi-Agent RAG via Internal Attention Rewards
Runsong Jia, Qian Chang, Mengjia Wu, Ciprian Doru Giurcaneanu, Jie Lu and Yi Zhang . 38648

GNMR: Runtime Stability Control for Low-Precision Large Language Model Training
Boao Kong, Weichen Jia, Engao Zhang, Guohong Li, Yonghan Dong, Yao Wang, Yaoyuan Wang, Yunke Peng and Kun Yuan 38665

Lost in Sampling: Assessing Lexical Reachability in LLMs via the Word Coverage Score (WCS)
Samer Awad, Javier Conde, Carlos Arriaga, Tairan Fu, Javier Coronado-Blázquez and Pedro Reviriego 38690

Data Shapes What, Scale Shapes How: Lexical Learning Dynamics in LLMs
Nazanin Shafiabadi and Guillaume Wisniewski 38703

RIDGE: Region-Informed Derivative-Guided Evidence Selection for Long Video Understanding
Shanqing Xu, Meng Luo, Mengchen Qian, Yuhui Gao, Siyue Peng, Xiaohan Zhong, Xiaojin Zhang, Zhongyu Wei, Wei Chen and Xiang Bai 38717

Reliability-Prioritized Fine-Grained Generation in Multimodal Large Language Models
Xiaomeng Fan, Wei Wu, Yuwei Wu, Zhi Gao, Luo Shiyu, Mingyang Gao, Haoyu Zhao, Zhenxin Diao, Yuxuan Ba, Lijia Feng, Yunde Jia and Mehrtash Harandi 38735

One Polluted Page Is Enough: Evaluating Web Content Pollution in LLM Recommenders
Minghao Luo and Liang Chen 38779

Rationale-Grounded In-Context Learning for Time Series Reasoning with Multimodal Large Language Models
Qingxiang Liu, Zhiqing Cui, Xiaoliang Luo, Yuqian Wu, Zhuoyang Jiang, Huaiyu Wan, Sheng Sun, Lvchun Wang, Wei Yu and Yuxuan Liang 38806

ECO-Mem: An Event-Centric Organization Framework for Visibility-Controlled Long-Term Memory
Xing Pei, Dandan Song, Zhijing Wu, Yuhang Tian, Jun Yang, Yitian Ma, Yading Li, Xudong Li, Huipeng Ma and Yan Xu 38825

Local Evidence, Global Credit: Visual Evidence-Calibrated Policy Optimization for Multimodal Reasoning
Zijing He, Zhongyi Deng, Kankan Lan, C.L.Philip Chen, Zeyu Chen, Enlei Gong and Tong Zhang 38844

From Bugs to Fixes: Self-Evolving Hierarchical Knowledge Graphs for LLM-Based Code Repair
Haoyue Ran, Junguo Zhu and Hongyu Ge 38859

MATE: Mobile Agent Trustworthy Execution for Safe and Efficient Mobile Control
Minsuk Jang, Hee-Seon Kim, Hui-Cheol Yu and Changick Kim 38874

From Repetition to Recognition: Inductive Discovery of Disinformation Narratives
Max Upravitelev, Veronika Solopova, Jing Yang, Charlott Jakob, Alexandra Tsiakalou, Neda Foroutan and Vera Schmitt 38896

Who Flips? Self- and Cross-Model Counterarguments Reveal Answer Instability in LLMs
Nafiseh Nikeghbal, Amir Hossein Kargaran, Shaghayegh Kolli and Jana Diesner 38928

CHRONOEXPERT: A Temporal Mixture-of-Experts for Knowledge Editing in Large Language Models
Xudong Li, Yizhou Jin, Dandan Song, Yongyu Huo, Yuhang Tian, Jun Yang, Chenhao Li, Huipeng Ma, Luan Zhang and Xing Pei 38943

SimRule: Collaborative Similarity-Driven Rule Learning for Interpretable Temporal Knowledge Graph Forecasting
Wenshuo Liu, Shufeng Hao, Ziyu Li and Chongyang Shi 38960

What Transfers Under Source Shift? Definitions, Examples, and Fine-Tuning for Climate Disclosure Classification
Guosheng Li, Fenghui Ren, Bin Liu, Chuan Yu, Kaiying Ji, Lin Yue, Jun Shen and Sasa Qian 38975

From Tokens to Topology: When Large Language Models Meet Graph-Structured Reasoning
Qihang Zhang and Liang Zeng 38990

Semantic Averaging with MUltiple Languages (SAMUL)
Cliff O'Reilly, Ernesto Jiménez-Ruiz and Tillman Weyde 39013

Contextualized Divergence Detection in Health Advice: A Benchmark and Analysis
Reet Yogesh Kothari, Sirajum Munir, Madhusudan Basak and Sarah Masud Preum 39035

Where Steering Signals Come From: Activation Source Selection in Activation Steering
Jiaran Ye, Lingxu Ran, Zijun Yao, Chenpeng Wang, Yong Jiang, Lei Hou, Juanzi Li and Liangming Pan 39061

Multi-Model Framework for Sentiment Annotation: A Use Case of Multilingual Wikipedia
Włodzimierz Lewoniewski, Milena Stróżyna and Izabela Czumałowska 39079

THE-Tree: Can Tracing Historical Evolution Enhance Scientific Verification and Reasoning?
Xin Wang, Jiyao Liu, Yulong Xiao, Junzhi Ning, Hairong Zhang, Lihao Liu, Junjun He, Botian Shi and Kaicheng Yu 39091

The Model Plants the Trigger: Answer-Side Backdoor Attacks in Multi-Turn Large Language Models
Yibo Zhang, Tianrong Guan, Liang Lin, Puze Wang, Jin Wang and Qingsong Wen 39113

Paved with True Intents: Intent-Aware Training Improves LLM Safety Classification Across Training Regimes
Jeremias Lino Ferrao, Niclas Müller-Hof, Iustin Sirbu, Traian Rebedea and Yftah Ziser . . . 39128

CoSteer: Collaborative Decoding-Time Personalization via Local Delta Steering
Hang Lv, Sheng Liang, Hao Wang, Hongchao Gu, Yaxiong Wu, Wei Guo, Defu Lian, Yong Liu and Enhong Chen 39156

HiP-LR: Hierarchical Planning with Localized Failure Recovery for Long-Horizon Multi-Robot Tasks
Chaerin Lee, Seoyoon Kim, Seungbeen Lee, Woojin Lee and Soohwan Song 39182

SICI: A Semantic-Pragmatic Complexity Index Reveals Regime Shifts in LLM Stance Detection
Fuqiang Niu and Bowen Zhang 39209

TADP: Structured Pruning of Large Language Models via Temporal Activation Dynamics
Adriano Puglisi, Flavia Monti, Francesco Leotta and Massimo Mecella 39222

Contradictory Evidence in LLMs: Tracing Internal Conflict Behavior with White-Box Signals
Passant Elchafei, Ali Monshizadehkamkar, Shahed Masoudian, Monorama Swain, Yufang Hou and Markus Schedl 39240

Exploring the Impact of Document Context on LLM-based Machine Translation Evaluation
Marina Iuliana Aur, Stefano Perrella and Roberto Navigli 39253

The Fragility of Jailbreak Robustness Across Operational States
Yuna Park, Hwang Youn Kim, Yujin Kim, Won Woo Ro, Suhyun Kim and Jae-In Hwang . 39277

DiaVLo: Diagnosing Behaviours of Vision-Language Models
Lorenzo Corti and Jie Yang 39299

MD-RE: A Multi-Discrimination Framework for Document-Level Relation Extraction with Adaptive Threshold Shifted Loss
Huangming Xu, Fu Zhang, Lu Zhang, Zhixuan Yang and Jingwei Cheng 39333

VideoNav: Cost-Aware Index Navigation for Long Video Understanding
Yulong Hui, Yuchen Li, Rui Kong, Xinran Chen, Jinman Zhao, Dawei Yin and Huanchen Zhang 39346

HalluPrism: When Multimodal Uncertainty Should Diagnose, Not Decide
Aman Prakash, Sourish Dasgupta and Tanmoy Chakraborty 39363

EvoFormat: Discovering Effective Graph-to-Text Formats via Rule-Based Evolutionary Search
Kazuya Kakizaki, Taiki Miyagawa, Hideyuki Oiso and Ryo Furukawa 39404

SemEnrich: Self-Supervised Semantic Enrichment of Radiology Reports for Vision-Language Learning
Halil Ibrahim Gulluk and Olivier Gevaert 39416

MemBook: A Hierarchical Reasoning-based Memory Framework for Hallucination-Resistant LLM Agents
Tongshu Bian, Yunxin li, Zhenran Xu, Tianqi Shi, Longyue Wang, Baotian Hu and Min Zhang 39444

ATSR: Adaptive Template Selection and Role-Aware Routing for Document-Level Multi-Event Argument Extraction
Yifan Wang, Tian Lan, Yanglei Gan, Run Lin, Leyuan Liu, Zufeng Wu, Xiaoyue Yan and Xiaohua Wu 39471

miniReranker: Efficient Multimodal Reranking through Visual Cache Reuse and Interaction Sparsity
Yingqi Fan, Xuan Lu, Anhao Zhao, Junlong Tong, Ping Nie, Kai Zou, Yunpu Ma, Wei Zhang and Xiaoyu Shen 39488

RULER: Instance-aware Rubric Rewards for SVG Generation
Hangyu Ran, Yuhao Zheng, Yingying Zhang, Kevin Qinghong Lin and Han Peng 39506

FourierSampler: Unlocking Non-Autoregressive Potential in Diffusion Language Models via Frequency-Guided Generation
Siyang He, Qiqi Wang, Xiaoran Liu, Hongnan Ma, Yiwei Shi, Yuerong Song, Ying Zhu, Tianyi Liang, Zengfeng Huang, Ziwei He and Xipeng Qiu 39525

Disentangled Privileged States for Simulator-Assisted Multi-Turn Dialogue Policy Training
Zhiyong Cao, Dunqiang Liu, Sicheng Zhou and Lin Cong 39543

Shared Low-rank Basis Factorization for Data-free Mixture-of-Experts Compression
Tianxiao Cao, Jiahe Shao, Yuning Qiu, Kyohei Atarashi, Hisashi Kashima and Qibin Zhao 39565

Does the Temporal Focus of Information Affect Confidence Calibration in LLMs? A Study of Temporal Effects on Model Confidence
Bhawna Piryani, Jamshid Mozafari and Adam Jatowt 39582

ASIL: Replacing Screenshot-and-Click with Structured State and Semantic Actions
Rui Xie and Lu Chen 39611

MCL: Meta-Collaborative Learning for Generalized Personalized Dialogue Generation
Muchen Yang, Chongming Gao, Fengbin Zhu, Weinan Gan, Wentao Shi, Jizhi Zhang, Quanyu Dai, Yong Liu, Yangyang Li and Fuli Feng 39630

One Policy, Any Budget: Internalizing Budget-Aware Search via Reinforcement Learning
Xiaowei Sun, Jin Li, Yili Hong, Yikun Fu and Yanghua Xiao 39643

How Correct Is Your Answer? A Semantic Correctness Framework for Open QA Evaluation
Elitsa Yotkova, Violeta Kastreva, Petar Velkov, Hristo Boyanov, Dimitar Iliyanov Dimitrov, Preslav Nakov and Ivan Koychev 39666

Token-to-Pixel Semantic Guidance for Document Image Tampering Localization
Tiancheng Wang, Haoyu Huang, Zichao Feng, Linlin Yang, Sheng Xu and Baochang Zhang 39686

TeleGen: Improving LLM-Based Web Application Generation via Runtime Telemetry
Yujia Luo, Haonan Zhang, Jiasi Shen, Zishuo Ding and Weiyi Shang 39697

Evaluating Inquiry Learning in LLM-Based Agents: Hypothesizing, Reasoning, and Exploration
Marta Knežević, Seyed Parsa Neshaei, Jibril Frej, Vinitra Swamy and Tanja Käser 39716

Agent-Trident: A Three-Level Progressive Diagnostic Benchmark for Tool-Use Hallucination in LLM Agents
Jiaqing Li, Du Xuehui, Liu Aodi, Jiaxi Shi and Jiangpeng Cui 39738

From Static Personal Values to Contextualized Personalization: Bayesian Personalized Value Alignment for LLMs
Hanze Guo, Aixuan Song, Jing Yao, Xiangxu Zhang, Xiaoyuan Yi, Xing Xie and Xiao Zhou 39762

CVeDRL: An Efficient Code Verifier via Difficulty-aware Reinforcement Learning
Ji Shi, Peiming Guo, Meishan Zhang, Miao Zhang, Xuebo Liu, Weili Guan and Min Zhang39791

DiPer: Disentangled Personal Direction Discovery for Personalized Generation Steering
Muchen Yang, Weinan Gan, Wentao Shi, Chongming Gao, Quanyu Dai, Yichao Wang, Yong Liu, Yangyang Li and Fuli Feng 39810

LLMZERO: Discovering Adaptive Training Strategies for RL Post-Training via LLM Agents
Haoyang Fang, Wei Zhu, Boran Han, Zhenyu Pan, Shuo Yang, Shuai Zhang, Jiading Gai, Peng Tang, Cuixiong Hu, Huzefa Rangwala, George Karypis and Bernie Wang 39827

Better Later Than Sooner: Neuro-Symbolic Knowledge Graph Construction via Ontology-grounded Post-extraction Correction
Lorenzo Loconte, Timothy Hospedales and Cristina Cornelio 39857

Block Recursive Transformers for Structured Parameter Sharing
Josep Marques, Hyunyoung A. Cho, Youngeun Kim, Youjia Zhang, Aecheon Jung, Kyuhong Shim and Sungeun Hong 39879

ProbeCourt: Cross-Examining Black-Box LLMs for Prompt-Injection Evidence Before Tool Execution
Jianjun Wu and Zexin Dong 39892

IndicHM: A Multilingual Benchmark and Disagreement-Centric Framework for Hateful Meme Detection
Shahid Shafi Dar, Mohd Aamir, Arindol Sarkar, Hrishiraj Chowdhury, Pranjal Pandey and Nagendra Kumar 39910

FreeMU: Free Marginal Utility for Reward-Aware Direct Preference Optimization
Lei Zhang, Jian Ma, Xianpeng Zhang, Dongyu Chen, Haonan Lu, Tianhuang Su, Chuangchuang Wang, Kai Tang and Zhixun Su 39932

Dynamic routing-based Mixture of Layer-wise Tokens with Token-Difficulty for Robust Audio-Visual Speech Recognition
Changhan Oh, Yeeun Jo, Ran Han, Jeomja Kang, Woo Yong Choi, Hwa Jeon Song and Kiyoung Park 39954

Beyond Single-Negative Preference: Multi-Negative DPO for LLM-Centric Historical Entity Linking
Tien Nam Nguyen, Emanuela Boros, Ahmed Hamdi, Adam Jatowt, Mickael Coustaty and Antoine Doucet 39970

Negative Advantage Is a Double-Edged Sword: Calibrating advantages in GRPO for Search Agents
Jiayi Wu, Ruobing Xie, Zeqian Huang, Lei Jiang, Can Xu, Kangyang Luo, Bochen Lin, Ming Gao and Xiang Li 39986

CulturalMenuBench: Probing the Knowledge-Application Gap in Multimodal Culinary Reasoning
Bo Zeng, Linfeng Gao, Peiqin Lin, Yu Zhao, Mingyan Zeng, Yu Tong, Xintong Wang, Linlong Xu, Longyue Wang, Weihua Luo, Qinggang Zhang and Jinsong Su 40003

Beyond Caption Loss: Generative Distillation for Modality Alignment in Vision-Language Models
Nam Kha Nguyen, Wei Emma Zhang, Xiang Dai, Mong Yuan Sim and Weitong Chen 40021

ARJUN: A Parameter-Efficient Multi-Objective Pre-Training Framework for Language Models
Subhadip Maji and Arnab Bhattacharya 40035

KVoiceBench, KOpenAudioBench, and KMMAU: Agent-Driven Korean Speech Benchmarks for Evaluating SpeechLMs
Haechan Kim, Seungjun Chung, Inkyu Park, Jihoo Lee and Jonghyun Lee 40053

Beyond Surface Forms: Symbolic Edits as a Test for Logical Reasoning with LLMs
Ramya Keerthy Thatikonda, Wray Buntine and Ehsan Shareghi 40070

Debiasing as a Measurement Intervention: Calibrated Ties and Resolution Loss in LLM-as-a-Judge Evaluation
Liang Zhao, Yong Wang and Jiangzhe Chen 40091

Video-DeepResearch: Towards the Next-Generation Multimodal Deepresearch Agent
Zhen Fang, Yu Zeng, Wenxuan Huang, Yiming Zhao, Shiting Huang, Tianfei Ren, Qi Lu, Qingnan Ren, Qisheng Su, Lionel Z. Wang, Qingyu Yin, Shuang Chen, Zehui Chen, Lin Chen, Zhenfei Yin, Yao Hu, Shaohui Lin, Wanli Ouyang, Shaosheng Cao and Feng Zhao 40113

UniCAR-RL: Seeing Better before Thinking Deeper in Visual Mathematics
Yuzhe Li, Hao Yan, Hao Wang, Xingchen Liu, Ya-Qi Yu, Jihao Wu, Minghui Liao, Wei Chen and Yuliang Liu 40132

Exploit More, Explore Smarter for Budget-Constrained Agentic Search
Haoyang Fang and Bernie Wang 40147

Can Audio LLMs Understand Prosodic Focus? An Alternative-Semantics Inference Test with Only"
Daniel Hardt 40166

S3C-LLM: Skill-Code Guided Agentic Language Models for Spectrum-to-Structure Elucidation
Xuanle Zhao, Xinyuan Cai, Xiang Cheng and Bo XU 40184

SOVER: Formal Certification of Optimization Reformulations via LLM-Assisted SMT Verification
Swapnil Bhattacharyya and Mayank Baranwal 40213

From Vision to Language: Investigating Causal Information Flow in Multimodal Decision-Making
Davide Testa, Hugh Mee Wong, Alessandro Lenci, Bernardo Magnini and Albert Gatt 40231

Probing for Long-Horizon Deductive Reasoning Capabilities in Language Models with Prolog
Hadeel Al-Negheimish, Jasna Ilieva and Yoon Kim 40248

Does Item Response Theory Generalize Across Languages? Evidence from Cross-Lingual LLM Benchmarks
Ruchira Dhar, Qiwei Peng and Anders Søgaard 40260

CS-CLIP: Compositional Scene Graph-guided CLIP for Robust Compositional Reasoning
SeongJun Jeong, Minjoon Jung, Woo Suk Choi, Youwon Jang and Byoung-Tak Zhang 40282

TokenRouter: Uncertainty-Guided Token-Level Routing for Efficient LLM Reasoning
Zhenning Shi, Yijia Zhu, Junhan Shi, Xun Zhang, Congcong Miao, Guorui Xie, Yong Jiang and Qing Li 40295

Anti-Prior Reinforcement Learning for Context-Grounded Reasoning
Zheng Wang, Zexin LU and Xuming Hu 40308

Efficient Frame Selection for Long Videos at Test Time with Attention-Based MLLM Selectors
Yilin Wang, Xiangxi Zheng, Dongxing Mao, Linjie Li, Zhengyuan Yang, Ping Yu, Rui Yan, Yuan Yao and Alex Jinpeng Wang 40328

DyBenchEval: A Comprehensive Evaluation Suite for Measuring LLM-as-Dynamic-Benchmarker
Pei Ke, Jingye Wang, Maolin Sun, Lan Wang, Haitao Peng, Bosi Wen, Enping Hu, Ruyi Luo, Minlie Huang and Ke Qin 40342

SEPO: Evidence-Grounded Prompt Optimization via Structural Editing
Xiaoyu Ma, Haoyue Liu, Yiwen li, Jionghao Zhu, Zhichao Wang, Ye Chen and Xiaoying Tang 40364

Beyond Binary Detection: A Multi-Dimensional Taxonomy of Cancer Misinformation on Reddit
Aria Pessianzadeh, Pooriya Jamie, Naima Sultana, Gracie Himmelstein, Yuliya Zektser, Patricia A. Ganz, Homa Hosseinmardi, Amir Ghasemian and Rezvaneh Rezapour 40392

GeoAgent: Evaluating VLM Geolocalization Through Embodied Navigation
Arka Mukherjee, Soham Roy, Kartikeya Trivedi and Shreya Ghosh 40420

MotionInsight: Diagnosing Object Motion Deficiencies in Generated Videos
Jiahao Zhan, Yongrui Ma, Qunliang Xing, Xuanyu Zhang, Jingqi Tong, Junlin Li, Li Zhang, Shijie Zhao and Tianfan Xue 40451

PLC-DPO: Posterior Label Correction in Noisy and Ambiguous Preference Optimization
Boryeong Cho, Sumyeong Ahn and Se-Young Yun 40467

Prompt-Robust Language Models: Which Training Strategies Work?
Frederic Sadrieh and Michal Štefánik 40490

GRIP: Gated Routing of Instance-level Preferences for Noisy Long-Tailed Text Classification
Ziqing Tian and Bo Yuan 40508

ICGR: Iterative Causal Graph Refinement for Document-Level Event Causality Identification
Longxuan Ma, Zhuoran Yu, Zhengtao Yu, Shengxiang Gao, Yuxin Huang and Yantuan Xian40536

Vectors from Larger Language Models Predict Human Reading Time and fMRI Data More Poorly when Dimensionality Expansion is Controlled
Yi-Chien Lin, Hongao Zhu and William Schuler 40556

SafeMath: Safe Solutions for Unsafe Math Word Problems
Sagnik Basu, Subhrajit Mitra, Aman Juneja, Somnath Banerjee, Rima Hazra and Animesh Mukherjee 40577

Synthetic Semantic Supervision for Contrastive Code Representation Learning in Small Transformers: An Empirical Study
Kenneth Paulsen, Florian Tambon, Mike Papadakis and Shin Yoo 40595

Func-R1: Incentivizing Mathematical Function Reasoning in Multimodal Large Language Models
Mingze Yin, Xiaohan Wang, Dian Li, Yao Haichao, Yilin Zhao, Youjun Chen, Sinbadliu, Jintai Chen, Yiheng Zhu, Chang-Yu Hsieh and Aimin Pan 40616

An Automated Pipeline for Provably Retrieval-Dependent Benchmark Construction over Dynamic Knowledge
Heng Zhou, Ao Yu, Yuchen Fan, Li Kang, Jianing Shi, Yongting Zhang, Yutao Fan, Tiancheng He, Yibing Lin, Hejia Geng, Yuhao Wu, Xiufeng Song, Zhemeng Zhang, Songtao Huang, Yiran Qin, Wenlong Zhang, Lei Bai and Zhenfei Yin 40633

Natural Language Defects in Code Generation Prompts: Empirical Impact and Lightweight Detection
Amal Akli, Mike Papadakis, Maxime Cordy and Yves Le Traon 40649

Same Tasks, Different Apps: Why Mobile GUI Agents Fail to Generalize?
Tien Tran, Nam Ho Koh, Daiki E. Matsunaga, Ayush Jain and Kee-Eung Kim 40664

More Context, Larger Models, or Moral Knowledge? A Systematic Study of Schwartz Value Detection in Political Texts
Víctor Yeste and Paolo Rosso 40694

Beyond Transfer Accuracy: Mechanism-Guided Controlled Adaptation for Low-Resource Languages
Khumaisa Nur'aini, Ayu Purwarianti, Alham Fikri Aji and Derry Tanti Wijaya 40715

Not All Tokens Are Equal: Syntax-Pattern Weighted Likelihood Difference Joint Detection for Neural-Training-Free AI-Generated Text Detection
Tianyu Guo, Shufeng Hao, JiaQi Li and Chongyang Shi 40743

GeoPrune: A Robust and Engine-Friendly Importance Signal Beyond Attention for Vision Token Pruning in VLMs
Kang Lou, Shijin Zhang, Ming Chen and Junzhe Cao 40759

LEGO: Synergizing Expert GraphRAG and Expert Chain-of-Thought for Legal Reasoning
Qingjing Chen, Junkai Zhang, Shaochun Wang, Jiahao Ding, Siyuan Zheng, Yukun Yan, Zhi Zheng, Antonino Rotolo, Yun Liu and Weixing Shen 40777

AWM: Answerable Working Memory for Long-Document VQA Agents
Dongzhuoran Zhou, Yuqicheng Zhu, Yule Liu, Zhen Yang, Rui Lu, Yuxiao Dong, Jie Tang and Evgeny Kharlamov 40824

What Makes Blood Thicker than Water: Related Word Discrimination in Romance Languages from a Multi-dimensional Perspective
Ana Sabina Uban, Teodor-George Marchitan, Ioan-Bogdan Iordache, Simona Georgescu and Liviu P Dinu 40840

SafeTutors: Benchmarking Pedagogical Safety in AI Tutoring Systems
Rima Hazra, Bikram Ghuku, Ilona Marchenko, Yaroslava Tokarieva, Sayan Layek, Somnath Banerjee, Julia Stoyanovich and Mykola Pechenizkiy 40864

Memory Beyond Recall: A Dual-Process Cognitive Memory System for Self-Evolving LLM Agents
Tianxiang Fei, Mingyang Song, Mao Zheng and Xiang Yu 40897

ToolHACE: A Dataset for Detecting Hallucinations in Tool-Based LLM Responses
Nikolay Ivanov, Sergey Pletenev, Elisei Rykov, Frolov Anatolii, Bogdan Monogov, Alexander Potekhin, Maria Marina, Daniil Moskovskiy, Viktor Moskvoretskii, Alexander Panchenko and Ilseyar Alimova 40915

Do Prompts Guarantee Safety? Mitigating Toxicity from LLM Generations through Gradient Based Subspace Intervention
Himanshu Singh, Ziwei Xu, A. Subramanyam and Mohan Kankanhalli 40935

Beyond Static Visual Tokens: Structured Sequential Visual Chain-of-Thought Reasoning
Guangfu Guo, Xiaoqian Lu, Yue Feng and Mingming Sun 40959

Understanding Subjectivity Transformation in Academic Meta-Review Generation
Sanju Basak and Suining He 40975

CePRM: Generative Process Reward Modeling with Calibrated Values and Repair-Aware Supervision
Zongxuan Li, Yixiang Dong, Bo Hu, Gonghao Cao, Yanzhen Chen and Huan Liu 40991

RULE-EDIT: Benchmarking Rule-Level Knowledge Editing in Large Language Models
Jizhi Zhang, Dylan Xinming Hou, Zhaoyi Li, Qishun Yang, Mahardika Krisna Ihsani, Muhammad Asif Ali, Gus Xia and Lijie Hu 41008

Pancasila-Dilemmas: Evaluating Large Language Models on Indonesian Human Value Dilemmas Grounded in Pancasila
Supryadi , Irfan, Julianti, Darren Keanly Martin, Jayvin Fernando, Yuqi Ren and Deyi Xiong 41028

ViCo: Towards Visual-oriented Coding with Self-Reflection for Chart Replication
Jiaxin Duan, Dian Jiao and Shuai Zhao 41043

Calibrating Hierarchical Memory Utilization for Long-Term Memory QA
Hang Lv, Xuyang Zhi, Hongchao Gu and Hao Wang 41072

Token-level Adaptive Weighting via Implicit Rewards for Inference-time Multi-objective Alignment
Dou Hu, Lingwei Wei, Peiran Xu, Hongjiang Xiao, Songlin Hu and Yuan Zhang 41082

T1-Bench: Benchmarking Multi-Scenario Agents in Real-World Domains
Genta Indra Winata, Amartya Chakraborty, Yuzhen Lin, Swasthi P Rao, Shikhhar Siingh, Houhan Lu, Nadia Bathaee, Sriharsha Hatwar, Paresh Dashore, Anmol Jain, Kshitij Tayal, Xiuzhu Lin, Anirban Das, Sambit Sahu and Shi-Xiong Zhang 41097

PEFT-Arena: Understanding Parameter-Efficient Finetuning from a Stability-Plasticity Perspective
Yangyi Huang, Ruotian Peng, Zeju Qiu, Jiale Kang, Yandong Wen, Bernhard Schölkopf and Weiyang Liu 41146

Learning from the Best, Differently: A Diversity-Driven Rethinking on Data Selection
Hongyi He, Xiao Liu, Zhenghao Lin, Mingni Tang, Yi Cheng, Jintao Wang, Wenjie Li, Peng Cheng and Yeyun Gong 41173

Aligning Large Language Models with Pluralistic Values: Case Study in Southeast Asian (SEA) Countries
Supryadi , Khue T. M. Luu, Maw Kyi Saw, Somchanmavong Souphanid, Yuqi Ren and Deyi Xiong 41196

Refinement-Integrated Reasoning Elicits Better Generalization of LLM-Based Code Vulnerability Detection
Xianda Du, Pei Ke, Enping Hu, Guannan Li, Jinyu Guo and Ke Qin 41215

BioPIE: A Biomedical Protocol Information Extraction Dataset for Experiment Understanding
Haofei Hou, Shunyi Zhao, Fanxu Meng, Kairui Yang, Lecheng Ruan and Qining Wang . . . 41234

Beyond Majority Vote: Multi-Perspective Adjudication for Medical Hallucination Detection
Joe Cecil and Marjorie Freedman 41258

K-Humor: Dynamic Multi-Agent LLM Framework for Generation of Culturally Grounded Korean Jokes
MinSu Park, Heejeong Jeon, DongJae Kim, Minwoo Oh and Eunil Park 41292

SWE-World: Building Software Engineering Agents in Docker-Free Environments
Shuang Sun, Huatong Song, Lisheng Huang, Jinhao Jiang, Ran Le, Zhihao Lv, Zongchao Chen, Yiwen Hu, Wenyang Luo, Xin Zhao, Hongteng Xu and Ji-Rong Wen 41316

One-Token Rollout: Guiding Supervised Fine-Tuning of LLMs with Policy Gradient
Rui Ming, Haoyuan Wu, Shoubo Hu, Zhuolun He and Bei Yu 41341

Towards Contextual Sensitive Data Detection
Liang Telkamp and Madelon Hulsebos 41354

RING: Retrieval-Internalized Generation for Continual Large-Scale Knowledge Injection
Shicheng Xu, Liang Pang, Liyi Chen, Zihao Wei, Jingcheng Deng, Yan Gao, Yi Wu, Yao Hu, Huawei Shen and Xueqi Cheng 41365

Exemplar Choice Matters for Morphological Reasoning: Model Confidence Vs Exemplar Similarity
Tsedeniya Kinfe Temesgen, Marion Di Marco and Alexander Fraser 41382

LLM generation novelty through the lens of semantic similarity
Philipp Davydov, Ameya Prabhu, Matthias Bethge, Elisa Nguyen and Seong Joon Oh 41411

Transferring Math Reasoning Abilities from LLMs to MLLMs via Causal-Aware Model Merging
Yijie Hu, Guanyu Yang, Zihao Zhou, Bin Dong, Xiaowei Huang, Kaizhu Huang and Qiufeng Wang 41432

Sparse Long Memory and Transition Bottlenecks Shape Positional Recall in Mamba Language Models
Muhammad Cendekia Airlangga, Hilal AlQuabeh, Munachiso Samuel Nwadike and Kentaro Inui 41448

TCHV: Type-Conditioned Hypothesis Verification for Multimodal Named Entity Recognition
Feiyu Han, Yang Liu, Yongxiu Xu, Yixiao Jiang, Zhaoqi Zhou and Chuan Zhou 41464

Toward Cultural Alignment: Human-Centered Evaluation of Multimodal AI Stories Across Five African Communities
Millicent Ochieng, Felermino D. M. A. Ali, Elizabeth A. Ankrah, Najeeb Gambo Abdulhamid, Migisha Boyd, Stephanie Nyairo, Mercy Muchai, Samuel Chege Maina, Aditya Vashistha, Anja Thieme and Jacki O'Neill 41484

Pivot Gradient Scaling: Reducing Dwell Time to Stabilize Extreme Low-Bit Quantization-aware Training
Vinnam Kim, Junhan Kim, Seungjae Shin, Jung Hyun Lee, Jaeseong You, An Chen and Jungwook Choi 41511

The Emergence of Relevance Through Axiomatic Attention Patterns During LoRA Fine-Tuning
Matthew Perlman, Atharva Nijasure and James Allan 41532

Probe and Steer: Localizing and Mitigating Position Bias in Pairwise LLM-as-a-Judge
Yulu Tang, Chun Kang and Yangfan Zhou 41549

PaveDSL: A Neuro-Symbolic Approach to Plan Specification and Verification for Multi-Agent Code Generation
Wutong Yu and Deyu Zhou 41566

LLMTrace: A Corpus for Classification and Fine-Grained Localization of AI-Written Text
Irina Tolstykh, Aleksandra Tsybina, Sergey Yakubson and Maksim Kuprashevich 41585

GATE: General Arabic Text Embedding for Enhanced Semantic Textual Similarity with Matryoshka Representation Learning and Hybrid Loss Training
Omer Nacar, Anis Koubaa, Serry Sibaee, Yasser Alhabashi, Adel Ammar and Wadii Boulila41623

Kwai Summary Attention
Chenglong Chu, Hao Peng, Hui Wang, Jiangxia Cao, Lu Ren, Zhiyuan Liang, Ziqi Wang, Hongtao Cheng, Yi Su, Xinchen Luo, Jian Liang, Ruiming Tang, Guorui Zhou and Han Li 41635

SaToMe: Salience-Driven Token Merging for Omni-Capable Diffusion MLLMs
Sixun Liu, Pengfei Guo, Jinghao Zhang, Tingyu Chen and Xiaofeng Zhang 41656

Skill Retrieval Augmentation for Agentic AI
Weihang Su, Jianming Long, Qingyao Ai, Qiaozhi He, Yichen Tang, Changyue Wang, Yiteng Tu, Yingbo Wang and Yiqun Liu 41673

Beyond Ranking: Role-Specific Signals for Token-Budgeted SFT Data Selection
Haocheng Zhu, Tianqi Wang and Jun Suzuki 41717

Influence Guided Sampling for Domain Adaptation of Text Retrievers
Meet Doshi, Vishwajeet Kumar, Yulong Li and Jaydeep Sen 41730

Beyond the Explicit: A Bilingual Dataset for Dehumanization Detection in Social Media
Dennis Assenmacher, Paloma Piot, Katarina Laken, David Jurgens and Claudia Wagner . . 41753

Efficient Agentic Reinforcement Learning with On-Policy Intrinsic Knowledge Boundary Enhancement
Dingwei Chen, Zefang Zong, Zhipeng ma, Leo Luo, Yang Li, Chengming Li, Peng Chen and Jie Jiang 41777

Values Are Not Enough: Situational Modifiers Shape Moral Decisions in LLMs and Humans
Peddi Manognya, Lokendra Mandloi, Joshi Sayali Shripad and Sandipan Dandapat 41796

DisasterLex: An Expert Concept-to-Schema Knowledge Graph for Geospatial Reasoning in Disaster Analytics
Yiming Xiao, Ankit Basu, Kai Yin, Sahil Rohit Vartak, Christian Swords and Ali Mostafavi 41818

Mitigating New-Tail Collapse in Long-Tailed Class-Incremental Learning with CLIP
HyeokJun Lee, Sungmin Cha and YoungJoon Yoo 41837

Where Do Apparent LLM Clinical Triage Failures Arise? Localizing the Multiple-Choice Format Effect
David Fraile Navarro, Berardino Como, Jialei Sheng, Soundariya Ananthan and Shlomo Berkovsky 41857

Beyond Binary: Turning Partial Success into Dense Verifiable Rewards for Reinforcement Learning in Code Generation
Longwen Wang, Yirui Liu, Xuaner Wu, Xiaohui Hu, Yuankai Fan, Kaidong Yu, Qizhen Weng, Wei Xi and Xuelong Li 41886

Binned Relative Policy Optimization towards Efficient Thinking of VLMs
ShuaiHang Sun, Zongzhao Li, Kai Shi, Jun Yang, Zhenyu Yang, Haonan Lu and Wenbing Huang 41903

S2J: Bridging the Gap Between Solving and Judging Ability in Generative Reward Models
Shaoning Sun, Jiachen Yu, Zongqi Wang, Xuewei Yang, Tianle Gu and Yujiu Yang 41919

Progressive Behavioral Drift through Compression Valleys in Large Language Models
Zhiyuan Xu, Stanislav Abaimov, Joseph Gardiner and Sana Belguith 41941

Towards Faithful and Controllable Personalization via Critique-Post-Edit Reinforcement Learning
Meiling Tao, Chenghao Zhu, Dongyi Ding, Tiannan Wang, Yuchen Eleanor Jiang and Wangchunshu Zhou 41965

TopoCurate: Modeling Interaction Topology for Tool-Use Agent Training
Jinluan Yang, Yuxin Liu, Zhengyu Chen, Chengcheng Han, Yueqing Sun, Xierui Song, Hui Su, Xunliang Cai, Fei Wu and Kun Kuang 41986

Selective Knowledge Edit Reversal via Gated Singular Vector Shrinkage
Weifeng Jiang, Ruirui Chen, Qianren Mao, Junnan Liu, Qili Zhang and Kwok-Yan Lam . . 42007

Text-to-SQL Benchmarks Toward Enterprise Realities: Under Massive Scopes, Complex Schemas and Scattered Knowledge
Yuchen Ji, Chengxi Yu, Jie Shi, Weixin Zhou, Xi Cao, Jiaqing Liang, Hai Chen, Yu Mao, Deqing Yang and Yanghua Xiao 42028

Full-Distribution Self-Distillation for Optimizing Soft Latent Reasoning
Ziyong Lin, Yanting Wang, Zixia Jia and Zilong Zheng 42061

Anom3W: A Unified Benchmark for Video Anomaly Grounding and Understanding
Liang Liang, Yun Zhu, Juekai Lin, Haoyang Wang, Jinhui Li, Weilin Zhou, Yuehan Zhu, Lijun Wu, Wenqiao Zhang and Jun Xiao 42074

Factorized Soft Prompts with Inference-Time Latent Control for LLM Personalization
Likun Yang, Pei Xu, Shiyue Cao, Tong Wang, Haoran Wu, Xiaotang Chen and Kaiqi Huang42093

Refinement Provenance Inference: Detecting LLM-Refined Training Prompts from Model Behavior
Bo Yin, Qi Li, Runpeng Yu and Xinchao Wang 42117

Surveying Representation Collapse in Pretrained Models of American Sign Language
Lee Kezar, Laurel Aichler, Athena S. Willis, Saki Imai, Mert Inan, Alicia K. Wooten, Malihe Alikhani and Lorna Quandt 42132

EYELIGNR: Aligning Gaze and Text via Contrastive Learning for Improved Reading Scanpath Prediction
Ekta Sood, Simon Tannert, Prajit Dhar, Meghdut Sengupta, Enrica Troiano and Sidney K. DMello 42147

HiDet: Hierarchical AI Text Detection via Coarse Filtering and Multi-Grained Contrastive Learning
Xiao Yang, Feifei Kou, Lei Shi, Yuhan Yao, Jiwei Zhang, Pengfei Zhang and Zhongbao Zhang 42161

Mind the Motions: Benchmarking Theory-of-Mind in Everyday Body Language
Seungbeen Lee, Jinhong Jeong, Donghyun Kim, Yejin Son and Youngjae Yu 42180

FAIR-Eval: Reliable and Efficient Adaptive Benchmarking for Large Language Models
Ziliang Pang, Tianhe Wang, Tianshu Yu, Ziyi Zhang, Shuo Ren and Jiajun Zhang 42196

RubricReviewer: From Direct Critique to Rubric-Driven Review via Agentic Evidence Synthesis and Human-Aligned Generation
Shuyu Guo, Wenxiang Hu, Yuyue Zhao, Yougang Lyu, Xiaohui Yan and Zhaochun Ren . . 42221

PersuaRL: Reinforcement Learning-Driven Multi-Expert Selection for Persuasive Dialogue Generation in Insurance
Rohan Kirti, Akash Ghosh, Aryan Vats, Niladri Ghosh, Shipra Shriparn, Roshni Ramnani, Anutosh Maitra and Sriparna Saha 42235

Filter by Quality, Stop by Difficulty: Confidence-Aware Self-Consistency with Adaptive Early Stopping
Jooyeong Na, Wansik Jo and Hyunsouk Cho 42269

Routing Before Looking: Query-Adaptive Evidence Acquisition for Long-form Video Understanding
Tianyue Wang, Xuying Wu, Yuxiang Ma, Ruiming Liang, Jiaxuan Kang, Yanchao Hao, Zheng Wei, Leigang Qu, Haiyun Guo and Jinqiao Wang 42290

Skill-CDPO: Evolving Agent Tool-Use via Critical Step Preference Optimization
Yuchen Li, Jinghao Lin and Yuanqing Wang 42309

Learning When to Think in Latent Visual Space: Reasoning Mode Adaptive Reinforcement Learning
Paul Jian, Jinpeng Ou, Leo Wei, Qingpei Guo and Jianing Li 42327

MMLDSum-LLM: Multimodal Long-Document Summarization with Visual-Alignment and Keyword-Aware
Xianpeng Zhang, Jiahua Yang, Dongyu Chen, Lei Zhang, Jian Ma, Guohuan Xu, Haonan Lu, Tianhuang Su, Chuangchuang Wang and Kai Tang 42350

Relational Over-Regularization: Graph-Based AI-Generated Text Detection via Sentence Transition Deviation
Hyeonchu Park and Bugeun Kim 42371

Directional Drift in Arithmetic Reasoning: How Comparative Cues Steer LLM Predictions
Mohammadamin Shafiei, Hamidreza Saffari and Nafise Sadat Moosavi 42392

CoReflect: A Reflective Co-Evolution Framework for Improving Conversational Evaluation
Yunzhe Li, Richie Yueqi Feng, Tianxin Wei and Chin-Chia Hsu 42407

Active Adaptation, Not Static Defense: Temporal Dynamics of Preventative Steering in Adversarial Fine-Tuning
Jing Guan, Yachao Yang, Zhaoliang Liu, Yuyao Zhang, Fanyu Meng and Junlan Feng 42426

The Benchmark Is Not the Construct: Toward Distribution-Based Evaluation in NLP
Haithem Afli and Manuel Frank 42441

When Charts Deceive Web Agents: A Paired Benchmark for Misleading Visualization-Grounded Tasks
Yisheng Li, Dongming Han, Hongkun Pan, Bingcheng Mao, Zijian Wang, Yunlong Mao, Wei Tong and Sheng Zhong 42466

Moral Safety in LLMs: Exposing Performative Compliance with Puzzled Cues
Mohammadamin Shafiei, Shuyue Stella Li and Yulia Tsvetkov 42485

Reliability-Calibrated Guidance for Few-Step Categorical Flow-Map Text Generation
Zihao Liu, Hongchen Wu, Xiaochang Fang, Hongxuan Li, Huaxiang Zhang and Jiande Sun42517

APEX-EM: Non-Parametric Online Learning for Autonomous Agents via Structured Procedural-Episodic Experience Replay
Pratyay Banerjee, Masud Moshtaghi and Ankit Chadha 42547

DisGAR: Machine Generated Text Detection via Disentangled Generative Artifact Representation
Shuhua Ma, Bohao Li, Lingcui Zhang, Yanru He, Ben Niu and Fenghua Li 42569

Guardrails without Shackles: Mitigating Over-Defense in Inference-Time Detection of Fixed-Target Backdoors for LLMs
Bingxu Han, Fang Wang, Bo Zhang, Xingyou Xia, Jun Zhang, Yunbo Cao, Zhunchen Luo, Xiaoying Bai, Guotong Geng and Mingqiang Wang 42584

Anchored by an Afterthought: Cross-Modal Numerical Anchoring in VLMs
Taehyeon Park, Jeonghun Park and Youngjae Yu 42608

Not All LoRA Parameters Are Essential: Insights on Inference Necessity
Guanhua Chen, Yutong Yao, Ci-Jun Gao, Lidia S. Chao, Yujia Huo, Feng Wan and Derek F. Wong 42635

Beyond Known Event Types: Open-Domain Event Detection via Ontology-Guided Grounding–Discovery Resolution
Yuhan Hu, Yupeng Qi, Yongxi Luo, Chen Su, Shan Huang and Ziyu Lyu 42649

Energy-Aware Benchmarking of Generative Information Extraction
Andrei Politov, Iryna Vergunova, Rene Jäkel and Michael Färber 42668

A Hyperbolicity Atlas of Large Language Model Hidden States
Zhichao Yang, Yuanze Hu, Gen Li, Qingchen Yu, Shiying Duan, Xinyu Wang, Ye Qiu, Zeming Liu, Guangxu Chen and Zhaoxin Fan 42692

Leak-Resistant Unlearning: A New Benchmark for Evaluating Multi-Hop Reasoning Consistency and Recovery Robustness
Haoting Qian, Qingjie Zhang, Zhicong Huang, Cheng Hong and Han Qiu 42702

MetaGen: Self-Evolving Roles and Topologies for Multi-Agent LLM Reasoning
Yimeng Wang, Jiaxing Zhao, Hongbin Xie, Hexing Ma, Yuzhen Lei, Shuangxue Liu, Xuan Song, Zichen Zhang and Haoran Zhang 42721

Stratified Topographic Similarity
Jean-Baptiste Sevestre, Matthieu Geist and Emmanuel Dupoux 42735

Padamitra: Grounded Glossary Generation for Classical Sanskrit
Manoj Balaji Jagadeeshan, Sai Pragnaan Marala and Pawan Goyal 42745

Routed Graph Handoff: Adaptive Format Selection for Multi-Agent LLM Delegation
Pratyay Banerjee and Ankit Chadha 42756

HiSVR: Hierarchical Semantic Visual Reconstruction for Unified Multimodal Models
Tailai Chen, Weining Wang, Yuchen Liu, Jun Fu, Jie Jiang, Yiming Cheng and Jing Liu . . . 42766

STAR: Detecting Inference-time Backdoors in LLM Reasoning via Prior-Posterior Amplification
Seong-Gyu Park, Sohee Park, Jisu Lee, Hyunsik Na and Daeseon Choi 42780

Sample Learning State Guided Dynamic SFT–RL Integration Post-training
Zheng Yang, Yilin Liu, Zhuoran Li, Wenhao Gao, Bingwei Li, Haozhe Sun, Caixia Yuan, Fan Wan, Wu Chen and Xiaojie Wang 42798

FrameBench: A Language Understanding Benchmark Based on Frame Semantics
Chihiro Yano and Ryohei Sasano 42819

Case-based Statute Retrieval for Multi-turn Legal Consultation with Fact-level and Rationale-level Mixed Evidence Graphs
Wanli Wu, Shizhu He, Yanchao Hao, Jun Zhao and Kang Liu 42834

MEDitor: A Plug-and-Play Medical Multi-Agent Routing Protocol
Yunwei Wang, Xianrui Zhang, Qiankun Ma, Rongjie Wang, Zhen Song and Haofei Wang . 42848

Tacit Coordination of Large Language Models
Ido Aharon, Emanuele La Malfa, Michael J. Wooldridge and Sarit Kraus 42863

Personalized News Headline Generation via Multi-Agent User Simulation
Li Kang, Yuhan Zhao and Li Chen 42900

DialPO: Dialectical Prompt Optimisation
Gokulakrishnan B, Vijay Kumar Neeluru and Vinayak S Puranik 42922

When Chain-of-Thought Fails, the Solution Hides in the Hidden States
Houman Mehrafarin, Amit Parekh and Ioannis Konstas 42939

The Role of Implicit and Explicit Demographic Signals in Large Language Model-based Student Assessment
Donya Rooein, Luca Benedetto and Dirk Hovy 42955

Beyond Static EHRs: Dynamic Examination Feedback for LLM-based Virtual Patients
Renwei Xiao, Wen He, Hongzhou Yu, Qingqiu Li, Yuejie Zhang, Rui Feng and Xiaobo Zhang 42980

Challenges and Recommendations for LLM-as-a-Judge in Multilingual Settings and for Low-Resource Languages
A. Seza Doğruöz, Xixian Liao, Verena Blaschke, Jakob Prange, Senyu Li and David Ifeoluwa Adelani 43001

VakyArth: Evaluating Pragmatic Competence in LLMs across Indic Languages
Usneek Singh, Poorvaja Veera Balaji Kumar, Parth Nanda, Anand Madhusoodanan, Geyang Guo, Wei Xu and Junyi Jessy Li 43027

Understanding Training Data Attribution in RLVR: From the Perspective of Causal Inference
Chi Zhang, Jian Zhao, Wei Wang, Kai Wang, Yao Zhao and Xuelong Li 43042

*"**Important** You should give me full credits!": Exploring Prompt Injection Attacks on LLM-Based Automatic Grading Systems*
Hang Li, Fedor Filippov, Yuping Lin, Pengfei He, Kaiqi Yang, Yucheng Chu, Yingqian Cui, Hui Liu and Jiliang Tang 43057

UniAttn: Reducing Inference Costs via Softmax Unification for Post-Training LLMs
Yizhe Xiong, Wei Huang, Xin Ye, Hui Chen, Zijia Lin, Haoran Lian, Zhenpeng Su, Jungong Han and Guiguang Ding 43072

ORDER: Task-Conditioned Routing for Retrieval-Augmented Generation
Aurélien Pellet, Marie Puren and Julien Perez 43092

EchoLens: A Human-Speech Dataset for Auditing Demographic Sensitivity in Audio-Language Models
Alejandro Salinas de Leon, Julian Nyarko and Allison Koenecke 43127

Reading the Geometry: A Mechanism for When Frozen LLM Text Helps Tabular Foundation Models
Zou Qingyun, Yingze Li and Nuo Chen 43179

Modeling Distinct Human Interaction in Web Agents
Faria Huq, Zora Zhiruo Wang, Zhanqiu Guo, Venu Arvind Arangarajan, Tianyue Ou, Frank F. Xu, Shuyan Zhou, Graham Neubig and Jeffrey P. Bigham 43195

Cross-lingual Functional Vectors for Emotion Detection in Large Language Models
Xue Jieying, Phuong Minh Nguyen, Le-Minh Nguyen and Shogo Okada 43213

Exploring Solution Divergence and Its Effect on Large Language Model Problem Solving
Hang Li, Kaiqi Yang, Yucheng Chu, Hui Liu and Jiliang Tang 43226

KODE: Mining Implicit Domain Knowledge from Historical Experience for Text-to-SQL
Qingshuo Lyu, Jiangyi Yin, Zhao Li, Meijie Du, Zhuojun Jiang and Qingyun Liu 43245

Phonologically Informed Tokenization for German Speech Recognition: A Cross-Domain Study
Christopher Witzl, Tobias Bocklet and Korbinian Riedhammer 43267

Vision-Default, Prior-Override: Causal Mechanisms of Perception-Knowledge Conflict in Vision-Language Models
Niclas Lietzow, Danielle Bitterman, Carsten Eickhoff, William Rudman and Michal Golovanevsky 43282

YallaMorph: A Benchmark for Evaluating Arabic Morphological Generation in Large Language Models
Mahmoud Reda, Salam Khalifa, Reham Marzouk and Nizar Habash 43299

Pigeonholing: how bad prompts hurt models, causing collapse and mistakes
HyunJi Nam, Keertana Chidambaram, Dorottya Demszky and Natasha Jaques 43316

ShizhenGPT: Towards Multimodal LLMs for Traditional Chinese Medicine
Junying Chen, Zhenyang Cai, Zhiheng Liu, Yunjin Yang, Rongsheng Wang, Qingying Xiao, Xiangyi Feng, Zhan Su, Jing Guo, Xiang Wan, Guangjun Yu, Haizhou Li and Benyou Wang . . . 43353

Relevance Supervision Can Teach Neural Retrievers to Prefer LLM-Generated Text
Wei Huang, Keping Bi, Yinqiong Cai, Wei Chen, Jiafeng Guo and Xueqi Cheng 43376

Iterative Natural Logic Verification for Multi-hop Question Answering
Jihao Shi, Liang Yang and Hongfei Lin 43407

Steering Vectors are an Adversarial Attack Surface
Abzal Aidakhmetov, Donato Crisostomi, Tommaso Mencattini, Adrian Robert Minut, Iacopo Masi and Emanuele Rodolà 43420

HTA: Temporal Reasoning Transformer with Temporal Context Encoding for Temporal Knowledge Graph Forecasting
Yingsong Ning, Fu Zhang, Jingwei Cheng, Jiashun Peng and Xiaoke Wang 43440

Cross-lingual Self-Consistency for Multilingual Reasoning with Language Models
Ahmed Elhady, Eneko Agirre and Mikel Artetxe 43454

How Many Instances Are Enough? Minimum Subsample Size Prediction for Stereotype Evaluation
Bora Bozdogan, Soroush Vosoughi and Weicheng Ma 43467

Unsupervised Post-Training of Foundation Models: A Survey
Yijie Xu, Qianyi Cai, Huizai Yao, Yili Wang, Tianfu Wang, Cehao Yang, Xingbo Yao, Zhiyu Guo, Aiwei Liu, Xuming Hu, Weiyu Guo and Hui Xiong 43485

From Tokens to Regions: CUDA-Sensitive Instruction Tuning for GPU Kernel Generation
Wentao Chen, Jiace Zhu, Qu Zeng, Qiaoling Xiao, Liucheng Duan and An Zou 43505

How Much Parallelism Is Free"? A Principle of Near-Free Parallelism for Parallel Decoding
Minghua He, Lingzhe Zhang, Yuan Liu, Zhou Xiao and Aiwei Liu 43520

EvA: An Evidence-First Audio Understanding Paradigm for LALMs
Xinyuan Xie, Shunian Chen, Zhiheng Liu, Yuhao Zhang, Zhiqiang Lv, Liyin Liang and Benyou Wang 43579

Critic-R: Improving Agentic Search using Instruction-tuned Retrievers with Natural Language Introspective Feedback
Md Zarif Ul Alam, Alireza Salemi and Hamed Zamani 43599

Interpretable Predictability-Based AI Text Detection: A Replication Study
Adam Skurla, Dominik Macko and Jakub Simko 43613

Notes to Self: Can LLMs Benefit from Experiential Abstractions?
Chang Liu, Xinyu Li and Artur Dubrawski 43634

Act More, Decide Less: Skill-Guided Adaptive Action Chunking for Long-Horizon LLM Agents
Yanting Yang, Can Jin, Jinman Zhao, Jiahao Wu, Yang Zhou, Zhepeng Wang, Zhendong Wang, Mu Zhou and Dimitris N. Metaxas 43646

Copying Before Suppression: What Drives a Below-Chance Dip During Language Model Training?
Tejas Dahiya and Cole Blondin 43668

K-FinHallu: A Hallucination Detection Benchmark for Multi-Turn RAG in Korean Finance
Eunbyeol Cho, Yunseung Lee, Mirae Kim, Jeewon Yang, Youngjun Kwak and Edward Choi 43718

FORGE: Forensic Reasoning with Grounded Evidence
Rohit Kundu, Shan Jia, Vishal Mohanty, Athula Balachandran and Amit Roy-Chowdhury . 43745

Bias-Aware Orchestration: Role-Specialized Multi-Agent Collaboration of Heterogeneous LLMs
Yuanpei Zhao, Haowei Zhang, Ziyue Qiao, Jiancheng Lv, Mao li and Wei Ju 43768

Precision over Scale: A Polish–Silesian Benchmark and a Translation System Outperforming Open-Source and Commercial Models
Grzegorz Kulik, Mikołaj Pokrywka, Adam Jatowt and Wojciech Kusa 43783

SciEmbed: Citation-Context Supervision for Scientific Document Embeddings
Jonas Wilinski and Michael Färber 43793

Can LLMs Design Computational Machines? Pushdown Automaton Synthesis as a Test of Structured Computational Reasoning
Syed Mumtahin Mahmud, Nazira Jesmin Lina, Shahriyar Zaman Ridoy, S. M. Muhtasimul Hasan and Md Mosaddek Khan 43820

When Lower Privileges Suffice: Investigating Over-Privileged Tool Selection in LLM Agents
Kaiyue Yang, Yuyan Bu, Jingwei Yi, Yuchi Wang, Biyu Zhou, Juntao Dai, Songlin Hu and Yaodong Yang 43846

A Benchmark for the Generation and Evaluation of Scientific Architecture Diagrams in HTML/SVG
Kyeongman Park, Kang-il Lee, Janghoon Han, Yerin Hwang, Taegwan Kang, Minwoo Lee, Changho Lee and Kyomin Jung 43868

What Attention Recalls and Recurrence Controls in Hybrid Language Models
Kirill Afendulev, Alexey Dontsov, Elena Tutubalina and Anton Korznikov 43888

Brain Translation: Multitask Sequence-to-Sequence Modeling for Intracortical Attempted Speech Decoding
Michał Olak, Tommaso Boccato and Matteo Ferrante 43901

Below the Noise Floor: Bimodal Seed Collapse and Distinct Failure Modes in Small-Model Knowledge Distillation
Dipto Sumit, Sakib Ul Haque and Farig Sadeque 43917

WRDO: Weighted Refusal Direction Optimization in Aligned Language Models
Grigoriy Evseev, Vladimir Solodkin, Dmitry Bylinkin, Daniil Medyakov, Vladimir Aletov and Aleksandr Beznosikov 43929

UTP-Bench: Uncertainty-aware Travel Planning Benchmark
Etcharla Revanth Rao, Priyanshu Karmakar, Shubhojit Mallick, Manish Gupta, Shreya Ghosh and Abhik Jana 43945

RiskGuard: Hypergraph-Based Pre-Execution Risk Control for Long-Horizon Tool Use
Yandong Yan, Zesen Cai, Daming Shi, Yifei He, Ruiting Dai and Yu Huang 43979

SimDiff-DNS: Simplex Diffusion with Dynamic Noise Scheduling for Temporal Knowledge Graph Reasoning
Xiaoke Wang, Hechun Yang, Fu Zhang, Yiwen Chi, Kai Li, Hongzhi Chen, Yingsong Ning, Jiashun Peng, Jingwei Cheng and Shuangshuang Gu 43998

ActTraitBench: Quantifying the Knowledge-Decision Gap in Large Language Models via Human-Grounded Behavioral Validation
Yutong Yang, Chenxi Miao, Weikang Li and Yunfang Wu 44018

AgenticTagger: Generating Structured Item Representation for Recommendation with LLM Agents
Zhouhang Xie, Bo Peng, Zhankui He, Ziqi Chen, Yizhao Han, Isabella Ye, Benjamin Coleman, Noveen Sachdeva, Fernando Pereira, Julian McAuley, Wang-Cheng Kang, Derek Zhiyuan Cheng, Beidou Wang and Randolph G. Brown 44031

Web Search Is Not Enough: LLMs Fail to Account for Paper Retraction Status
Huy Hoang Ha, Benoit Favre, Cyril Labbe and François Portet 44052

Towards Understanding What State Space Models Learn About Code
Jiali Wu, Shweta Verma, Abhinav Anand and Mira Mezini 44078

A Systematic Survey of Large Language Model Routing
Chang Jiangeng, Xingyu Wu, KC Tan and Mengling Feng 44102

AgSynth: A Multi-Agent Reverse Synthesis Framework for Multimodal Mathematical Reasoning Data Generation
Yaxuan Wang, Lingfeng Zeng, Jiatu Hu, Guojie Lin, Boyu Yang, Bing Zhao, Lin Qu and hu Wei 44135

Lusopol: How do LLMs Handle Politically Charged Questions? The Use Case of Portugal
Giuseppe Attanasio, Inês Vieira, Inês Calvo, Beatriz Canaverde, Ben Peters, Orfeas Menis Mastromichalakis, Miguel Faria, James Furtado, Iago Paulo, David Semedo, Joao Magalhaes and Andre Martins 44149

UniFinEval: Towards Unified Evaluation of Financial Multimodal Models across Text, Images and Videos
Fangqi Lou, Zhi Yang, Lingfeng Zeng, Zhiheng Jin, Qi Qi, Wei Zhang, Zhenyu Wu, Zhenxiong Yu, Lejie Zhang, Xiaoming Huang, Xiaolong Liang, Zheng Wei, Junbo Zou, Dongpo Cheng, Zhaowei Liu, Xin Guo, Rongjunchen Zhang and Liwen Zhang 44168

Memes-as-Replies: Can Models Select Humorous Manga Panel Responses?
Ryosuke Kohita and Seiichiro Yoshioka 44207

TRIPPULSE: Multi-Agent Travel Planning with Review-Grounded Reasoning
Priyanshu Karmakar, Borru Vijay Sai, Shubhojit Mallick, Abhik Jana, Shreya Ghosh and Manish Gupta 44223

When Disagreement Meets Noise: Noise Robust Annotator Embeddings for Subjective NLP
Kevin Jacob, Parsa Ghazvinian, Uthman Jinadu, Nicole Tianjiao Yang and Yi Ding 44254

CPSD: Counterfactual Patches Self-Distillation in RLVR
Yixiao Zhou, Dongzhou Cheng, Hehe Fan and Yu Cheng 44272

SMUGRI-COPA: Evaluating Commonsense Reasoning for Low-Resource Finnic Languages
Taido Purason, Marili Tomingas, Janek Vaab and Mark Fishel 44289

SinoGlyphBench: A Diagnostic Benchmark for Chinese Glyph-Level Obfuscation in Language-Model Moderation
Yifan Wang, Zimu Wang, Suliu Qin, Changyu Zeng, Tong Chen, Siqi Chen, Yijie Lin, Lingyu Jiang, Jionglong Su, Yushan Pan, Haiyang Zhang, Wei Wang and Qiaoyu Tan 44304

MemesOverTime: A Decade-Scale Study of Meme Dynamics
Shijia Zhou, Diego Frassinelli and Barbara Plank 44328

Contextual Observer Grounding: Evaluating Situated Spatial Reasoning in Vision-Language Models
Mimo Shirasaka, Haochen Zhang and Yonatan Bisk 44344

TEVI: Text-Conditioned Editing of Visual Representations via Sparse Autoencoders for Improved Vision-Language Alignment
Sweta Mahajan, Sukrut Rao, Jiahao Xie, Alexander Koller and Bernt Schiele 44358

Do LLMs Use Cultural Knowledge Without Being Told? A Multilingual Evaluation of Implicit Pragmatic Adaptation
Mehwish Nasim, Sanjeevan Selvaganapathy, Neel Ganapathi Sabhahit, Marie Griesbach, Pranav Bhandari, Janina Lütke Stockdiek, Lennart Schäpermeier, Usman Naseem and Christian Grimme44383

Entropy Is Not Enough: Unlocking Effective Reinforcement Learning for Visual Reasoning via Vision-Anchored Token Selection
Senjie Jin, Peixin Wang, Boyang Liu, Xiaoran Fan, Shuo Li, Zhiheng Xi, Jiazheng Zhang, Yuhao Zhou, Tao Gui, Qi Zhang and Xuanjing Huang 44403

Riemannian Large Language Modeling
Shu XU, Jiahong Liu, Wenhao Zhuang, Jiabao li, Yefan Wang, Ming Sun and Menglin Yang44427

The Privacy-Hallucination Tradeoff in Differentially Private Language Models
Krithika Ramesh, Krishna Pillutla, Danish Pruthi and Anjalie Field 44441

Beyond Fooling: Model Manipulation Under Explanation-aware Training
Tabio Romanski and Stjepan Picek 44490

Intervention-Certified Amplification: A Validity Gate for σ-Weighted Post-Training
Leiyue Chen, Meng Wei, Xuefeng Hu and Wei Liu 44510

Don't Forget Your Embeddings: Robust Knowledge Erasure via Precise Editing of Embeddings
Clara Haya Suslik, Or David Shafran and Mor Geva 44523

Understanding Why Language Models Hallucinate: Testing Reasoning Against Priors
Yangfan Hu, Xuhan Tong, Haoyue Bai, Xi Ding, Shashank Muralidhar Bharadwaj, Siyang Cao, Robert D Nowak and Jiawei Zhang 44555

Judging with Confidence: Calibrating Autoraters to Preference Distributions
Zhuohang Li, Xiaowei Li, Chengyu Huang, Guowang Li, Katayoon Goshvadi, Bo Dai, Dale Schuurmans, Xin Zhou, Hamid Palangi, Yiwen Song, Palash Goyal, Murat Kantarcioglu, Bradley A. Malin and Yuan Xue 44582

Asclepius: An Adaptive Harness for Long-Horizon Clinical Agents
Grace Chang Yuan, Xiaoman Zhang, Sung Eun Kim, Luyang Luo and Pranav Rajpurkar . . 44611

Asymmetric Multimodal Learning with Residual Decomposition for Multimodal Intent Recognition
Yuxuan Zhai, Zongwei Li and Longfei Liu 44630

Reinforcement Learning Elicits Contextual Learning of Unseen Language Translation
Hanxu Hu, Zdenek Snajdr, Pinzhen Chen, Jannis Vamvas and Rico Sennrich 44643

A Recipe for Long-Context Reasoning in Large Language Models via On-Policy Optimization and Distillation
Miguel Moura Ramos, Duarte Miguel Alves and Andre Martins 44658

The Judge Can Debias Itself: Awareness and Self-Debiasing for Reliable LLM-as-a-Judge
Youbing Yin, Jingyu Wu, Anoop Kumar, Scott Novotney, Daben Liu and Erin Babinsky . . 44677

PercepLog-0: LLM-Augmented Backward Chaining for Legal Reasoning
Parisa Kordjamshidi, Sai Amrit Bulusu, Madhavan Seshadri, Samer Aslan, Leslie Barrett and Enrico Santus 44688

Beyond Retrieval: Learning Compact User Representations for Scalable LLM Personalization
Heng Cao, Fan Zhang, Jian Yao, Yujie Zheng, Changlin Zhao, Lu Hao, Yuxuan Wei, Wangze Ni, Huaiyu Fu, Yuqian Sun and Xuyan Mo 44707

DSL-LLaDA: Scaling Continuous Denoising to 8B Masked Diffusion LMs
Longxuan Yu, Yunshu Wu, Yu Fu, Siheng Xiong, Rob Brekelmans, Hui Liu, Yue Dong and Greg Ver Steeg 44727

An Algorithmic Account of Few-Shot Concept Induction in Language Models
Michael A. Lepori, Laila Johnston, Aalok Sathe, Zhuonan Yang, Roman Feiman and Ellie Pavlick 44753

Localizing Prompt Ambiguity in Large Language Models with Probe-Targeted Attribution
Govind Ramesh, Yao Dou and Wei Xu 44774

ERASE: Precise Machine Unlearning via Geometry-Aware Gradient Shaping and Adaptive Loss Control
Itzik Waizman, Shahar Katz and Lior Wolf 44796

From Distribution to Decision: A Diagnostic for Composable PTQ Primitives
Dorsa Zeinali and Yun Fu 44825

MoraLink: Bridging Morals and Narrative Fables for Retrieval
Lior Livyatan, Kai Golan Hashiloni, Asif Amar, Roei Rahamim and Kfir Bar 44855

Inference-Time Scaling in Diffusion Models: A Beam Search Approach to Visual Sequences Generation
Guilherme Fernandes, Vasco Ramos, Regev Cohen, Idan Szpektor and Joao Magalhaes . . . 44886

PragReST: Self-Reinforcing Counterfactual Reasoning for Pragmatic Language Understanding
Jihyung Park, Minchao Huang, Liu Leqi and Elias Stengel-Eskin 44910

VeriSim: A Configurable Framework for Stress-Testing Medical AI Under Patient Communication Noise
Sina Mansouri, Mohit Marvania, Vibhavari Shihorkar, Han Ngoc Tran, Kazhal Shafiei, Mehrdad Fazli, Yikuan Li and Ziwei Zhu 44933

Right Frame, Wrong Rule: Cultural Cues Expose the Financial Knowledge Gap They Were Meant to Close
Rania Elbadry, Ahmed Heakl, Saeed Almheiri, Fan Zhang, Muhra AlMahri, Xueqing Peng, Mohsinul Kabir, Shuyao Wang, Yi Han, Saadeldine Eletter, Duzhen Zhang, Preslav Nakov, Yuxia Wang, Fajri Koto and Zhuohan Xie 44959

AREAs-Lab: An Interactive Environment for AI-driven Requirement Elicitation for AI Systems
Pengshan Cai, Zihao Zhang, Ting Jin, Chenyang Zhu, Kushal Chawla, Sangwoo Cho, Scott Novotney, Yebowen Hu, Fei Liu, Shi-Xiong Zhang and Sambit Sahu 44987

Counterexample Generation via Per-Theorem Symbolic Verifiers: When Imitation Hurts and Reinforcement Repairs
Omar Farouk Zouak, Houssam Eddine Boukhalfa, Lakehal Soumaya, Shiv Katiyar and Prof Samia Nefti-Meziani 45032

ERGA: An Explorer–Refiner Game with Backdoor-Adjusted Verdict Aggregation for LLM-based Claim Verification
Wanlong Yu, Feng Liu, Zhiliang Tian, Zhen Huang, Yonghui Wang, Jingyuan Huang and Dongsheng Li 45048

Identifying Crucial Attention Heads for Multilingual Language Models: Retrieval and Retrieval-Transition Heads
Shaswat P Patel, Vishvesh Trivedi, Yue Han, Yihuai Hong and Eunsol Choi 45067

Separating Syntax from Language: A Mechanistic Account of Translation in Multilingual LLMs
Mikhail Sonkin, Tanja Baeumel, Daniil Gurgurov, Josef Van Genabith and Simon Ostermann 45092

Structural Collapse of Safety-Critical Weight Protection Scorers Under Group-Wise Post-Training Quantization
Jianchang Su and Wei Zhang 45113

A Multi-Agent Approach for Claim Verification from Tabular Data Documents
Rudra Ranajee Saha, Laks V. S. Lakshmanan and Raymond T. Ng 45135

Morphological Analysis and Disambiguation of Agglutinative Assamese through Rule-Based Synthesis and Neural Candidate Ranking
Pragyanur Saikia, Utpal Sharma and Arindam Karmakar 45155

When Chatbots Accommodate: Auditing the Response Policies of AI Companions in Vulnerable Conversations
Minh Duc Chu, Yifan Wu, Zhiyi Chen, Angel Hsing-Chi Hwang and Luca Luceri 45168

Balanced Aggregation: Understanding and Fixing Aggregation Bias in GRPO
Zhiyuan Zeng, Jiameng Huang, Zhangyue Yin, Bingrui Li, Jiashuo Liu, Yuhao Wu, Yining Zheng, Ge Zhang, Wenhao Huang, Xipeng Qiu and Ziniu Li 45195

PACEShop: Evaluating Personalized, Actionable, Compositional, and Evidence-grounded Shopping Assistants
Weimin Lyu, Chen Luo, Guangrui Li, Yaochen Xie, Dhineshkumar Ramasubbu, Arief Koesdwiady, Wanqiu Long, Hansu Gu, Yutong Chen, Zheshen Wang, Dakuo Wang and Yi Liu 45208

TRIS: A Tri-Layer Retrieval Integrity Sieve Against Knowledge Poisoning
Muhaimin Bin Munir, Akib Jawad Ononto, Nazia Shehnaz Joynab, Bhavani Thuraisingham and Latifur Khan 45250

PPE-Bench: A Benchmark for Evaluating MLLM Unlearning under Private-Public Entanglement
Xianren Zhang, Delvin Ce Zhang, Dongwon Lee and Suhang Wang 45265

TRACER: Trace-Regularized Training Against Indirect Prompt Injection
Kshitij Mishra, Salem Lahlou and Nils Lukas 45279

TreeFlash: Parallel AR-Approximation for Faster Speculative Decoding
Peer Rheinboldt, Frédéric Berdoz and Roger Wattenhofer 45297

PhysVidBench: Language-Grounded Evaluation of Physical Commonsense in Text-to-Video Models
Enes Sanli, Baris Sarper Tezcan, Erkut Erdem and Aykut Erdem 45311

The Hard Part Comes After Search: Benchmarking Web Agents on Synthesizing, Organizing, and Displaying Knowledge
Alexander Gill, Md Farhan Ishmam, Xuyen Nguyen, Neha Bhat, Parker Henry DeYoung, Fateme Hashemi Chaleshtori, Nathan Stringham, Kenneth Marino and Ana Marasovic 45360

Accurate but Not Humble: Evaluating Epistemic Humility in LLM Agents under Knowledge Conflict
Kaiser Sun, Bernal Jiménez Gutiérrez, Hongjun Liu, Jingyu Zhang, Jie Gao, Mark Dredze and Daniel Khashabi 45401

Many Dialects, Many Languages, One Cultural Lens: Evaluating Multilingual VLMs for Bengali Culture Understanding Across Historically Linked Languages and Regional Dialects
Nurul Labib Sayeedi, Md. Faiyaz Abdullah Sayeedi, Shubhashis Roy Dipta, Mahbub E Sobhani, Rubaya Tabassum, Ariful Ekraj Hridoy, Mehraj Mahmood, Md. Tarek Hasan and Swakkhar Shatabda 45421

Listening to the Slides: Multimodal LLM Speech Inpainting with Acoustic-Aware-Loss
Yanir Edri and Bracha Laufer-Goldshtein 45455

AutoData: Agentic Search for Pre-training Data Selection
Yan Meng, Dhruv Srikanth, Bingchen Zhao, Zhengyao Jiang and Yuxiang Wu 45474

When Can You Prune Your Network? A Study of Intermediate Neurons in Multilingual Speech Parsing
Minnie Kabra, Benjamin Lecouteux and Maximin Coavoux 45492

MediX-R1: Open Ended Medical Reinforcement Learning
Sahal Shaji Mullappilly, Mohammed Irfan Kurpath, Jaseel Muhammad Kaithakkodan, Omair Mohamed, Mohamed Zidan, Fahad Shahbaz Khan, Salman Khan, Rao Muhammad Anwer and Hisham Cholakkal 45507

Definitional Sensitivity in Media Bias Detection: A Multi-Definition Dataset and Benchmark
Martin Paul Wessel, Timo Spinde, Jürgen Pfeffer and Gianluca Demartini 45530

IndicSafe: A Benchmark for Evaluating Multilingual LLM Safety in South Asia
Priyaranjan Pattnayak, Garima Panwar and Sanchari Chowdhuri 45548

GRACE-SVD: Guided Rank Allocation and Core Equalization SVD
Andra Siva Sai Teja, Pragya Paramita Sahu, Vibhor Dave, Joshua Mahadevan, Ankit Mahato, Debanjan Konar and Vikram Mupparthi 45569

TaoAnalysisBench: Do Automated Theorem Prover LLMs Generalize Beyond MathLib?
Alexander K Taylor, Junyi Zhang, Ethan Ji, Vigyan Sahai, Haikang Deng, Yuanzhou Chen, Yifan Yuan, Di Wu, Jia-Chen Gu, Kai-Wei Chang, Nanyun Peng, Amit Sahai and Wei Wang 45588

AI translation of literary texts is 'fine', but readers still prefer human translations
Yves Ferstler, Adam Podoxin, Ty Brassington, Gaëlle Laperrière, Roman Grundkiewicz, Marie-Jean Meurs, Maite Taboada and Marzena Karpinska 45607

XLGoBench: Detecting cross-lingual skill gaps with algorithmic tasks
Purvam Jain, Vihari Piratla, Preethi Jyothi and Suvrat Raju 45660

Balancing Faithfulness and Performance in Reasoning via Multi-Listener Soft Execution
Nithin Sivakumaran, Shoubin Yu, Hyunji Lee, Yue Zhang, Ali Payani, Mohit Bansal and Elias Stengel-Eskin 45711

InterviewBench: Benchmarking Large Language Models as Interviewers
Shangzhe Dai, Feiyu Duan and Zhongyu Wei 45731

Detecting Post-generation Edits to Watermarked LLM Outputs via Error-Correcting Codes
Zewei Deng, Muhammad Siddeek, Liyan Xie, Mohamed Mohamed, Mengdi Wang, H. Vincent Poor and Andrea Goldsmith 45788

MoEs Are Stronger than You Think: Hyper-Parallel Inference Scaling with RoEs
Soheil Zibakhsh Shabgahi, Mohammad Samragh, Kumari Nishu, Lauren Hannah, Arnav Kundu and Minsik Cho 45797

Multi-Hop Knowledge Composition is Bound by Pretraining Exposure
Yannis Karmim, Luis Martí, Djamé Seddah and Valentin Barriere 45815

FinRT: Distilling Adaptive Red-Teaming Strategies into Reusable Adversarial Generators in Consumer Finance
Rikhiya Ghosh, Himanshu Kumar, Sriram Venkatapathy, Sahil Wadhwa, Alexandre G.R. Day and Pranab Mohanty 45826

Han2Han: Efficient Language-Specific Character Representation through Script-Aware Pre-Training for Historical Text Analysis
Cellik Adams, EunKyoung Jo and Juae Kim 45857

A History-Aware Visually Grounded Critic for Computer Use Agents
Jaewoo Lee, Zaid Khan, Archiki Prasad, Justin Chen, Supriyo Chakraborty, Kartik Balasubramaniam, Sambit Sahu, Elias Stengel-Eskin, Hyunji Lee and Mohit Bansal 45877

Health-SCORE: Improving Open-Ended Health LLMs Using Scalable Rubric-Based Rewards
Zhichao Yang, Sepehr Janghorbani, Dongxu Zhang, Jun Han, Qian Qian, Andrew Ressler II, Gregory D Lyng, Sanjit Singh Batra and Robert E. Tillman 45909

What Really Controls Temporal Reasoning in Large Language Models: Tokenisation or Representation of Time?
Gagan Bhatia, Ahmad Muhammad Isa, Maxime Peyrard and Wei Zhao 45931

OASIS: Online Anchor-calibrated Self-Improvement System
Wenyi Wang, Firas Laakom, Li Nanbo, Yanning Dai, Yuhui Wang, Yifan Xu, Dmitrii Khizbullin, Piotr Piękos, Yilan Zhang, Yimeng Chen, Alfonso Amayuelas, Mingchen Zhuge and Jürgen Schmidhuber 45950

Language Modeling Across Script Variants: The Case of Japanese
Go Inoue, Nizar Habash and Timothy Baldwin 45964

AMARIS: A Memory-Augmented Rubric Improvement System for Rubric-Based Reinforcement Learning
Peilin Wu, Xinlu Zhang, Kun Wan, Wentian Zhao, Gang Wu, Xinya Du and Zhiyu Chen . . 45981

Sense and Sensitivity: Benchmarking LLM Clinical Triage Recommendations with Physician Experts
Abinitha Gourabathina, Haoran Zhang, Yuexing Hao, Walter Gerych and Marzyeh Ghassemi 46053

Behavior Carries Over: An AI Sandbagging Detection Method Inspired by Honest Inertia
Junjie Chu, Mingjie Li, Zeyuan Chen, Ye Leng, Yisen Wang, Michael Backes and Yang Zhang 46083

When Trivia Is Not Trivial: Everyday Knowledge Failures in Multilingual LLMs
Anna Mosolova and Djamé Seddah 46100

DP-IPI: A Hybrid Differential Privacy Text Rewriting Mechanism for Indirect Personal Identifiers in Clinical Texts
Ibrahim Baroud, Stephen Meisenbacher, Sebastian Möller, Florian Matthes and Roland Roller 46129

Training with Subword Variants Delays Overfitting in Repeated-Data Language Modeling
Adrian Cosma, Stefan Ruseti, Emilian Radoi and Mihai Dascalu 46144

Don't "Overthink" Pointwise Reranking: Is Reasoning Truly Necessary?
Nour Jedidi, Yung-Sung Chuang, James R. Glass and Jimmy Lin 46159

Beyond Lexical and Syntactic: Better Text Simplicity Analysis Informed by Psycholinguistic Research
Benjamin Vendeville, Liana Ermakova and Pierre De Loor 46173

BEiTScore++: Reference-free Image Captioning Evaluation with BEiT-3 and Conformalized Quantile Regression
Goncalo Emanuel Cavaco Gomes, Bruno Martins and Chrysoula Zerva 46192

RFCLLM: Evaluating LLMs' Reasoning Ability of Network Protocol State Machines
Anqi Chen, Dan Goldwasser and Cristina Nita-Rotaru 46227

Characterizing Opinion Evolution of Networked LLMs
Caleb Probine, Yigit Ege Bayiz, Filippos Fotiadis, Samuel Li, Yunhao Yang and Ufuk Topcu 46247

From Arraignment to Resolution: Transformer Models of Court Cases
Kory Rosen and Aron Culotta 46272

Gnosis Deepsearch: Structured Search Space Navigation for Complex QA
Umut Halil, Le Zhang, Jin Huang, Yuwei Zhang, Xinyu Hou, Andre Melo, Shriram Piramanayagam, Amela Fejza, Ruofei Lai, Yang Ren, Pavlos Vougiouklis and Jeff Z. Pan 46292

Mind the Phrase: Annotation and Identification of Multiword Expressions in Grammatical Error Correction
Shiva Taslimipoor, Christopher Bryant, Zheng Yuan, Diane Nicholls, Andrew Caines and Paula Buttery 46316

CultureScore: Evaluating Cultural Faithfulness in Video Generation Models
Anku Rani, Wei Dai, Shravan Nayak, Patricia Maes, Mahdi M. Kalayeh and Paul Pu Liang46330

Gradient-Stable Attention Heads Signal LLM Correctness
Sophie Ostmeier, Brian Axelrod, Maya Varma, Asad Aali, Yabin Zhang, Magdalini Paschali, Sanmi Koyejo, Curtis Langlotz and Akshay S Chaudhari 46364

UI-Oceanus: Scaling GUI Agents with Synthetic Environmental Dynamics
Mengzhou Wu, Yuzhe Guo, Yuan Cao, Haochuan Lu, Dfshenzhu, Pingzhe Qu, Xin Chen, Kangqin, Zhongpu Wang, Xiaode Zhang, Xinyi Wang, Wei Dai, Gang Cao, Yuetang Deng, Gong Zhi, Dezhi Ran, Linyi Li, Wei Yang and Tao Xie 46381

MCompassRAG: Topic Metadata as a Semantic Compass for Paragraph-Level Retrieval
Amirhossein Abaskohi, Raymond Li, Gaetano Cimino, Peter West, Giuseppe Carenini and Issam H. Laradji 46400

Spatial Reasoning via Modality Switching Between Language and Symbolic Representations
Shreya Rajpal, Tanawan Premsri and Parisa Kordjamshidi 46423

Humans' ALMANAC: A Human Collaboration Dataset of Action-Level Mental Model ANnotations for Agent Collaboration
Jiaju Chen, Yuxuan Lu, Jiayi Su, Chaoran Chen, Songlin Xiao, Zheng Zhang, Yun Wang, Yunyao Li, Jian Zhao, Tongshuang Wu, Toby Jia-Jun Li, Dakuo Wang and Bingsheng Yao 46459

Holmes-Graph: A Framework for Probing the Linguistic Competence of Graph-Based Text Embeddings
Andric Valdez Valenzuela, Selene Baez Santamaria and Helena Gomez Adorno 46482

Debug or Regenerate? Test-Guided Selection for LLM Code Repair
Atharv Kulkarni, Archiki Prasad, Vivek Srikumar, Elias Stengel-Eskin, Hyunji Lee and Mohit Bansal 46494

EMBRACE: Emotion-Based Representation Across Cultures & Expressions
Pravin Mote and Carlos Busso 46509

SproutRAG: Attention-Guided Tree Search with Progressive Embeddings for Long-Document RAG
Amirhossein Abaskohi, Issam H. Laradji, Peter West and Giuseppe Carenini 46521

How Calibration Content Shapes Attention-Based Reranking
Petros Karypis, Hossein Rajaby Faghihi, Peter Chen, Rui Zhu, Noveen Sachdeva, Yan Zhu and Julian McAuley 46536

Do LLMs Choose Like Humans? Using Cognitive Theory to Evaluate LLM Decision-Making
Johnathan Sun, Andrei Shleifer and Yonatan Belinkov 46552

Latent Visual Sensitive Fingerprinting for Black-Box Vision–Language Model Tamper Detection
Chaoxiang He, Jichen Chai, Hanqing Hu, Yi Wang, Shuo Wang, Hongsheng Hu and Bin Benjamin Zhu 46570

A Dataset for Modeling Iterative Problem-Solving
Fagun Patel, Sang T. Truong, Duc Quang Nguyen, Kazunori Fukuhara, Nick Haber, Ben Domingue and Sanmi Koyejo 46590

ReVision: Scaling Computer-Use Agents via Temporal Visual Redundancy Reduction
Amirhossein Abaskohi, Yuhang He, Peter West, Giuseppe Carenini, Pranit Chawla and Vibhav Vineet 46621

PersonaForge: Realistic Multi-Turn User Simulation for Agentic Systems
Hanglong Lv, Dawei Zhu, Lei Li, Bowen Ye, Huaqiu Liu, Yifan Song, Bofei Gao, Weimin Xiong, Jinhao Dong, Chenhong He, Lingpeng Kong, Qi Liu, Tong Yang and Fuli Luo 46647

EVA-Bench: A New End-to-end Framework for Evaluating Voice Agents
Tara Bogavelli, Gabrielle Gauthier Melancon, Katrina Stankiewicz, Oluwanifemi Bamgbose, Fanny Riols, Hoang H Nguyen, Raghav Mehndiratta, Lindsay Devon Brin, Hari Subramani, Joseph Marinier, Sridhar Krishna Nemala, Anil Madamala and Srinivas Sunkara 46671

Prompt Optimization with Human Feedback
Xiaoqiang Lin, Arun Verma, Zhongxiang Dai, See-Kiong Ng, Patrick Jaillet and Bryan Kian Hsiang Low 46749

Reinforced Internal-External Knowledge Synergistic Reasoning for Efficient Adaptive Search Agent
Ziyang Huang, Xiaowei Yuan, Yiming Ju, Haoran Cai, Jun Zhao and Kang Liu 46768

CAROB: A Benchmark for Cross-Artifact and Authority-aware Reasoning over Organizational Bundles
Eunbin Kim, Hyuk Namgoong, Wooyoung Go, Seungjin Ryu and Sangkeun Jung 46786

Do Backdoored LLMs Share Internal Trigger Representations? Evidence from Frozen SAE Feature Banks
Minjeong Choi, Jaesin Ahn and Heechul Jung 46805

Conditioning Activation Interventions on Evolving Reasoning States in Large Language Models
Tae-Won Kim, Hee-Dong Kim, Jumin Han and Seong-Whan Lee 46826

Markovian Generation Chains in Large Language Models
Mingmeng Geng, Amr Mohamed, Guokan Shang, Michalis Vazirgiannis and Thierry Poibeau 46847

Not Safe for All: Auditing the Dialect Penalty in Text-to-Image Safety Pipelines
Minkyu Kim, Juhwan Choi and YoungBin Kim 46866

TEX: Test-Time Scaling Testing Agents via Execution-based Cross-Validation
Akash Gokul, Srijan Bansal, Yingbo Zhou and Semih Yavuz 46889

Entropy Meets Importance: A Unified Head Importance–Entropy Score for Stable and Efficient Transformer Pruning
Minsik Choi, Hyegang Son, Changhoon Kim and Young Geun Kim 46895

Do We Require Large Language Models To Be HHH Aligned For Reasoning, Safety, Value, and Pluralistic?
Gautam Siddharth Kashyap, Mark Dras and Usman Naseem 46936

C-QUERI: Congressional Questions, Exchanges, and Responses in Institutions Dataset
Manjari Rudra, Daniel Magleby and Sujoy Sikdar 46951

Benchmarking Speech-to-Speech Translation Models
Alkis Koudounas, Hayato Futami, Quentin Jodelet, Osamu Take, Shinji Watanabe and Emiru Tsunoo 46973

VisualThink-VLA: Visual Intermediate Reasoning for Effective and Low-Latency Vision-Language-Action Policies
Mingjian Gao, Wenqiao Zhang, Yuqian Yuan, Yang Dai, Binhe Yu, Zheqi Lv, Haoyu Zheng, Jiaqi Zhu, Zhiqi Ge, Zixuan Wan, Siliang Tang and Yueting Zhuang 47009

MRMAD: A Multi-Round Multi-Audio Benchmark for Evaluating Acoustic Degradation Perception in Large Audio-Language Models
Yize Li, Ningyuan Yang, Sile Yin, Sindhuja Thogarrati, Sung-En Chang, Andrew C. Singer, Xue Lin, Chuan-Che Huang and Shuo Zhang 47028

3D Visual Prompting: Empowering 3D-Aware Visual Generation via In-Context Self-Refinement
Reina Ishikawa, Fumiaki Sato, Ryo Hachiuma, Taiki Sekii and Hideo Saito 47055

Control-Data Flow Separation: Stable Prompt Optimization in Multi-Agent LLMs
Wentao Zhang, Syed Shariyar Murtaza, Junaid Ahmad Bhatti, Utkarsh Soni, Yifan Nie, Eugene Wen and Yuntian Deng 47076

Can Omni-models Imagine the Future? A Diagnostic Benchmark for Multimodal World Modeling
Junyeob Kim, Jihwan Kim, Minha Jhang, Sungwook Jeon, Hwiyeol Jo, Jeonghoon Kim, Sanggoo Lee and Joosung Lee 47099

Toward Autonomous Long-Horizon Engineering for ML Research
Guoxin Chen, Jie Chen, Lei Chen, Zhao Jiale, Fanzhe Meng, Xin Zhao, Ruihua Song, Cheng Chen, Ji-Rong Wen and Kai Jia 47116

OPERA-MoE: Output-aware Routing Perturbation Alignment via Observable Discrepancy for MoE Quantization
Geonho Lee, Hancheol Park, Tae-Ho Kim and Jungwook Choi 47132

AutoLexSteer: Automatic Contrast Construction for Lexical Activation Steering
Shuhe Wang, Lachlan Cowley, Eduard Hovy and Jey Han Lau 47149

Knowledge Acquisition During Pre-training? Large Language Models Learn Better With Auxiliary Views
Joseph Lee, Yidi Huang, Dokyoon Kim, Shu Yang and Li Shen 47175

Discovery, Articulation and Analysis of Communication Framings
Rakshitha Rao Ailneni and Sanda M. Harabagiu 47198

OpenSeal: Good, Fast, and Cheap Construction of an Open-Source Southeast Asian LLM via Parallel Data
Tan Sang Nguyen, Muhammad Reza Qorib and Hwee Tou Ng 47243

Depth as Verifiable Geometric Feedback: Enhancing Spatial Reasoning in Multimodal Large Language Models
Yepeng Zhang, Ruohang Xu, Zihao Yue, Boshen Xu and Qin Jin 47258

The Interplay of Harness Design and Post-training in LLM Agents
Kyungmin Kim, Youngbin Choi, Seoyeon Lee, Suhyeon Jun, Dongwoo Kim and Sangdon Park 47275

Proactive Safety Alignment: Enabling Proactive Initiative in Multi-Turn Defence
Mengyuan Qin and Yue Feng 47300

MetaCLASS: Metacognitive Coaching for Learning with Adaptive Self-regulation Support
Naiming Liu, Richard Baraniuk and Shashank Sonkar 47312

EMR-AGENT: An LLM Agent Framework for Cohort and Feature Extraction from Heterogeneous EHR Databases
Kwanhyung Lee, Sungsoo Hong, Joonhyung Park, Jeonghyeop Lim, Juhwan Choi, Donghwee Yoon and Eunho Yang 47341

Revisiting Complete Reasoning Traces for Post-Training
Jaehui Hwang, Sangdoo Yun, Byeongho Heo and Dongyoon Han 47374

Auditing Speaker-Level Metadata Utility in Text-Residual Prosody Prediction
Gaeun Kim, Jihyun Park, Sohee Jang and Joon-Hyuk Chang 47393

Understanding and Enhancing Backdoor Persistency in LLM Agent Post-Training
Qiusi Zhan, Nian Lyu, Stephanie Ding, Arnav Mehta, Xander Davies and Daniel Kang . . . 47413

Inside Sentence Embeddings: Leveraging Token Representations for Machine Translation Evaluation and Word Alignment
Mana Makinae, Takumi Goto, Yusuke Sakai, Hidetaka Kamigaito and Taro Watanabe 47431

Do Spoken Language Models Hear Speech as They Read Text? Bridging Structural Gaps Between Speech and Text
Hyeonyu Kim, Hwayeon Kim, Youngwon Choi, Myeongkyun Cho and Huu-Kim Nguyen 47439

Dynamic Hub-and-Spoke Memory for Streaming Video Understanding
Xinru Jiang, Lin Zhao, Xi Xiao, Yunbei Zhang, Janet Wang, Chenrui Ma, Haolin Li, Yanzhi Wang, Yifan Gong and Octavia Camps 47455

DTRNet: Dynamic Token Routing Network to Reduce Quadratic Costs in Transformers
Aman Sharma, Saeed Najafi, Parsa Farinneya, Benyamin Jamialahmadi, Marzieh S. Tahaei, Yuhe Fan, Mehdi Rezagholizadeh, Boxing Chen and Aref Jafari 47476

Not Worth Another Token: Marginal Value Estimation for Efficient Deep Research Agents
Harshitha Kolukuluru, Reshma Ashok, Kirat Arora, Evan William Ciccarelli, Nischal Ashok Kumar, Lunyiu Nie, Franck Dernoncourt, Samyadeep Basu, Ryan A. Rossi and Nedim Lipka 47490

Token-Disentangled Latent Test-Time Scaling for Vision-Language Reasoning
Hao-Xuan Ma, Yihao Liu, Yutao Sun, Yanting Miao, Mengyu Zhou, YiCheng Xiao, Long Chen, Zhenguo Li, Han-Jia Ye, Xiaoxi Jiang and Guanjun Jiang 47517

Modality Disentangled Learning for Incomplete Multimodal Emotion Recognition: A Primitive Memory Distillation Perspective
Jiaqi Zhang, Zheng Pang, Mengting Li, Yiqi Wang, Guangyuan Dong, Chao Xue, Yusen Wu, Zihao Li, Huy Phan, Sicheng Zhao, Björn Schuller and Jiachen Luo 47534

Self-Evolving World Models for LLM Agent Planning
Xuan Zhang, Wenxuan Zhang, See-Kiong Ng and Yang Deng 47553

The Bottleneck Is Retrieval, Not Reading: Diagnosing Evidence-Based Generation with RAG for Long Context Question Answering
Debopriyo Banerjee, Owais Aijaz, Ahmer Tabassum, Zhuohan Xie, Preslav Nakov and Veselin Stoyanov 47578

STRESS-FG: Diagnosing Visual Grounding in Zero-Shot Fine-Grained Recognition
Ishtiaque Ahmed Khan and A. Lynn Abbott 47614

When to Change the Canvas? Safe Adaptive Variable-Length Decoding for Diffusion Language Models
Junzhuo Li, Tianhao Li, Hao-Ran Wei, Xiang Li, Baosong Yang and Xuming Hu 47641

Multi-Label Delegation Target Identification in Japanese Law
Tsuyoshi Fujita, Yuya Sawada, Yusuke Sakai and Taro Watanabe 47655

Supporting and performing culture from the inside
Lea Frermann and Steven Bird 47683

From Empathy to Personalized Empathy: Adapting Empathetic Strategies to Individual Users
Wuqiang Zheng, Chengbing Wang, Yilin Yang, Junyi Cheng, Jianfei Xiao, Sun HU, Yi Xie, Yangyang Li and Wenjie Wang 47692

PV-Decoding: Progressive Verification Decoding with Decode Gating for Diffusion Language Models
Renxi Liu, Yiming Li, Shengyin Sun, Chen Chen, Xian Wang, Shuai Yang, Chen xi, Wenbo Gao, Xiaoguang Li, Hui-Ling Zhen, Xianzhi Yu, Weizhe Lin, Xiaosong Li and Yaoyuan Wang 47710

DeepPersona: Scaling Persona Depth via Taxonomy-Guided Generation
Zhen Wang, Yufan Zhou, Zhongyan Luo, Lyumanshan Ye, Adam Wood, Saab Mansour, Man Yao and Luoshang Pan 47721

Listwise Explanation of Embedding-Based Rankings via Semantic Chunk Grouping
Hyunkyu Kim, Yeeun Yoo and Youngjun Kwak 47743

Accelerate Speculative Decoding with Sparse Computation in Verification
Jikai Wang, Jianchao Tan, Yuxuan Hu, Jiayu Qin, Yerui Sun, Yuchen Xie, Xunliang Cai, Juntao Li and Min Zhang 47762

From Decoding Knobs to Commit Frontiers: Understanding Trajectory Control in Discrete Diffusion Language Models
Junzhuo Li, Bo Wang, Tianhao Li, Hao-Ran Wei, Baosong Yang and Xuming Hu 47776

Causally Redundant Reasoning: A Dual-Disruption Test Framework for Measuring Shortcut Tendencies in Chain-of-Thought
Jiwoo Kim and Jongwuk Lee 47788

MAS-Bai: Linguistically Informed Adaptive Multi-Agent Translation for Bowen-Script Bai
Jingting Zheng, Hong Ren, Yuehan Cui, Shaolin Zhu and Deyi Xiong 47821

ChronoClinQA: A Longitudinal Multi-Admission Benchmark for Clinical Reasoning over Electronic Health Records
Tong Chen, Zimu Wang, Dongzhe Fan, Luoyao Chen, Haoyi Lian, Yifan Wang, Zhengni Yang, Zhengyong Jiang, Procheta Sen, Wei Wang, Jionglong Su and Qiaoyu Tan 47835

BackendForge: Benchmarking Agentic End-to-End Code Generation with Backend Services
Yuzhe Guo, Mengzhou Wu, Yuan Cao, Jialei Wei, Dezhi Ran, Wei Yang and Tao Xie 47861

OriGround: Orientation-Aware Neuro-Symbolic Zero-Shot 3D Visual Grounding
Haochen Li, Jiaxin Shi, Ruonan Liu and Luo Liufu 47877

Aligning LLMs for Crystal Material Design: Test-Time Reinforcement Learning and Model Merging
Pengyu Wang, Yang Han, Xin Chen, Kai Yu, Tong Zhu, Lu Chen and Bo Chen 47898

Beyond Refusal: Constructive Communication Rewriting for Harmful User Language
Moxin Li, Jingyang Qi, Wynne Hsu and Mong-Li Lee 47920

Diagnosing Knowledge Conflict in Multimodal Long-Chain Reasoning
Jing Tang, Kun Wang, Haolang Lu, Hongjin Chen, KaiTao Chen, Zhongxiang Sun, Qiankun Li, Lingjuan Lyu, Guoshun Nan and Zhigang Zeng 47947

ABSOL: Aggregated Bayesian Subsampling Orchestrated with LLMs
Jackson Hassell, Chen Shen and Estevam Hruschka 47987

Before the First Token: Jailbreak Detection with Attention Probes
Pedro Faustini, Annabelle McIver and Mark Dras 48011

KARMA: Karma-Aligned Reward Model Adaptation
Jared Scott and Jesse Roberts 48030

ZeoAgent: A Closed-Loop Compiler for Coupled OSDA–Condition Generation in Zeolite Synthesis
Zhongan Wang, Zhaolin Hu, Kun Li, Zhiliang wu, Ming Li and Hehe Fan 48040

How Much Retrieval Depth Does Your Reader Need? Noise Ceilings and Cost-Effectiveness in Table QA
Jheng-Peng Huang, Yao-Chung Fan and Jiun-Long Huang 48057

Sparsity in Linear Attention Models: Bregman Testing Time Learning
Wenqi Zeng, YuAng Chen, Yuxuan Chen, Chumin Sun, Weihuang Wen, Yichuan Liu, Li Zhou, Tian Wang, Fan Zhang, Yuan Yao and Jie Sun 48075

Beyond Output Consistency: Diagnosing Representation-Level Causal Controllability in LLM Persona Simulation
Chaoyong Jiang, Shengling Wang, Ke Chao, Tianyu Hao, Kunpeng Jiang and Zhaolin Yang48095

TopicMerge: A Divide and Conquer Approach to Topic Modeling
Matthew Raffel, Lizhong Chen, Sai Munikoti and Sameera Horawalavithana 48120

MUSE: Multi-Domain Chinese User Simulation via Self-Evolving Profiles and Rubric-Guided Alignment
Zihao Liu, Hantao Zhou, Jiguo Li, Jun Xu, Jiuchong Gao, Jinghua Hao, Renqing He and Peng Wang 48136

Ask or Answer: A Decision Framework for Multi-Turn Health Misinformation Intervention
Xiaoying Song, Anirban Saha Anik, Jinyu Liu, Qitao Tan, Geng Yuan and Lingzi Hong . . . 48149

FinanceHarness: Autonomous Financial Deep Research Framework
Yijia Xiao, Rujun Han, Yanfei Chen, Zifeng Wang, Ke Jiang, Zhongying CuiZhu, Wei Wang, Tomas Pfister and Chen-Yu Lee 48172

Generate What You Can Trust: Content Credibility in Generative Recommenders
Zhuo Cai, Guanghao Wu, Shoujin Wang, Peilin Zhou and Victor W. Chu 48198

F-GraphRAG: Preserving Fine-Grained Facts in Graph Retrieval-Augmented Generation
Shenmeng Chen and Kang Zhang 48219

Visual Search Augmented Chain-of-Thought Reasoning for Attribute Value Extraction from Product Videos
Tong Wu, Ming Cheng, Jiazhen Hu, Jiaying Gong and Hoda Eldardiry 48238

AdaC-GRPO: Towards Mastering Long-Horizon Visual Sudoku Puzzles via Adaptive Curriculum Learning
Sida Li, Yuanxin Liu, Linli Yao, Shicheng Li, Xu Sun, Haoli Bai and Xiao-Hui Li 48255

Understanding Moral Reasoning Trajectories in Large Language Models: Toward Probing-Based Explainability
Fan Huang, Haewoon Kwak and Jisun An 48275

Evaluating Second-Order Bias of LLMs Through Epistemic Entitlement
Ramaravind Kommiya Mothilal, Terry Jingchen Zhang, Raiyan Ahmed, Zhijing Jin, Shion Guha and Syed Ishtiaque Ahmed 48323

OpenFinGym: A Verifiable, Multi-Task Gym Environment for Evaluating Quant Agents
Kaicheng Zhang, Wen Ge, Lei Jiang, Weixin Yang, Jordan Langham-Lopez, Jialin Yu, Lukasz Szpruch and Hao Ni 48348

Characterizing Narrative Content in Web-scale LLM Pretraining Data
Teagan Johnson, Elliott Ash, Andrew Piper and Maria Antoniak 48364

SKILL-DISCO: Distilling and Compiling Agent Traces into Reusable Procedural Skills
Zhongxin Guo, Danrui Qi, Hanwen Gu, Peng Cheng and Yongqiang Xiong 48405

†DAGGER: Distractor-Aware Graph Generation for Executable Reasoning in Math Problems
Zabir Al Nazi, Shubhashis Roy Dipta and Sudipta Kar 48423

Enhancing Low-Resource Language Reasoning via High-Resource Language Feature Transfer
Minju Song, Hyeon Hwang, Junhyun Lee and Jaewoo Kang 48447

Before the Arrest: Benchmarking LLMs on Criminal Profiling from Incomplete Evidence
Yutong Yao, Cao Yanjie, Guanhua Chen, Yang Xu, Junchao Wu, Zeyu Wu, Lidia S. Chao and Derek F. Wong 48462

Quality, Not Just Outcome: A Structural Scoring Framework for Selecting Coding-Agent Trajectories for Supervised Fine-Tuning
Sida Li, Yuanxin Liu, Shicheng Li, Linli Yao, Xu Sun, Haoli Bai and Xiao-Hui Li 48484

FinanceLLM: A Survey of Large Language Models in Finance
Ojas Nimase, Zhengao Li, Xinyu Wei, Jinglin Hu, Luojia Liu, Kangyi Zhao, Yue Zhao and Yushun Dong 48501

CONDUIT: A Unified Residual-Stream Restoration Framework for KV Cache Reuse in Vision-Language Models
Pengan Chen, Kaisheng Zheng, Liang Hong, Lixia Yi, Jiyue Jiang, Jiayang Chen, Yixuan Wang, Yimin Fan, Xinyuan Liu, Jiayi Li, Zhanqiu Zhang, Steven Y. Guo and Yu Li 48515

MATCHA: A Chinese Multi-Disorder Mental Health Dataset with Cross-Lingual Analyses on Social Media
Bingrui Jin, Baihan Li, Zitong Li, Shimeng Li, Kunyao Lan, Kenny Q. Zhu and Mengyue Wu 48536

ReasonFlow: Flow Matching that Thinks Step by Step
Bing Yan, Ge Sun, Juan J. de Pablo, Stefano Martiniani and Kyunghyun Cho 48554

TRILOGUE: A Trilingual Spoken Dialogue Fact-Checking Benchmark with Evidence and Paired Audio
Chaewan Chun, Meruyert Aristombayeva, Jiyoung Choi, Mahjabin Nahar, Delvin Ce Zhang and Dongwon Lee 48572

StreamAlign: Streaming Text-Aligned Speech Tokenization
Kang-wook Kim, Jinyoung Park, Jinsoo Kim, Sehun Lee, Tony Woo and Gunhee Kim 48596

The Plot Twist: Jailbreaking Unified Multimodal Models with a Three-Act Narrative Attack
Shaoxiong Guo, Tianyi Du, Lijun Li, Yuyao Wu, Jie Li and Jing Shao 48616

Dialects of Translationese Shape Language Model Learning
Jenny Kunz 48640

EvoRubrics: Dynamic Rubrics as Rewards via Adversarial Co-Evolution for LLM Reinforcement Learning
Hongxin Ding, Baixiang Huang, Yue Fang, Weibin Liao, Zheng Li, Jinyang Zhang, Zhijing Wu, Junfeng Zhao and Yasha Wang 48660

When Bias Scores Reflect Abstention Failure: Disentangling Stereotype Direction from Refusal to Abstain in BBQ
Jeemin Kim and Jooyoung Kim 48681

Learning to Act Beyond Language: Expanded Action Spaces for Sequential Decision-Making with LLMs
Zhongqi Yue, Weishi Wang, Yundaichuan Zhan, Juncheng Li, Daniel Dahlmeier and Fredrik D. Johansson 48698

Guard Models Are Overconfident Where Base Models Are Uncertain
Jonghyun Hong, Min Jae Jung and Minwoo Kim 48726

Online Strategic Reasoning for LLM-Based Negotiation via Opponent-Aware Monte Carlo Tree Search
Chen-Cheng Huang, Yi-Cheng Wang, Hung-Lin Hu, Yi-Ren Yeh and Chu-song Chen 48740

Instruction-Tuned Language Models Cannot Sample from Distributions They Can Describe
Chaemin Jang, Jihee Kim and Dongman Lee 48757

Who Do Language Models Think Is Competent? A Mechanistic Analysis of Occupational Bias
Keren Fuentes and Aaron Mueller 48778

MediSketch: Test-Time Scaling for Multimodal Patient Education
Zhipeng Tang, Benlu Wang, Zhenyu Zhang, Wuao Liu, Pengda Xiang, Juncheng Huang, Chin Siang Ong, Hong yu and Zonghai Yao 48808

Latent Recurrent Transformer: Architecture Exploration, Training Strategies, and Scaling Behavior
Zeyi Huang, Xuehai He, Liliang Ren, Yiping Wang, Baolin Peng, Hao Cheng, Shuohang Wang, Pengcheng He, Jianfeng Gao, Yong Jae Lee and Yelong Shen 48860

Expectations and Practices around AI Disclosure in CS Research
Arati Mohapatra and Danish Pruthi 48880

APEX: Academic Poster Editing Agentic Expert
Chengxin Shi, Qinnan Cai, Zeyuan Chen, Long Zeng, Yibo Zhao, Jing Yu, Jianxiang Yu and Xiang Li 48904

Towards Training Transformers as a Universal Computer
Ruize Xu, Chenxiao Yang, Yanhong Li and David McAllester 48933

HypoExplore: Agentic Discovery with Active Hypothesis Exploration
Jaywon Koo, Jefferson Hernandez, Ruozhen He, Hanjie Chen, Chen Wei and Vicente Ordonez 48955

ExpCSD: Experience-based Debiasing Framework for Conversational Stance Detection Grounded in Latent Thinking Paths
Xiang Zeng, Ye Wang, Xin Song, Dong Zhu, Wenqing Hou and Bin Zhou 48998

Do Gender Cues Affect LLM Value Trade-offs? Evidence from a Controlled Decision Benchmark
Yangyang Liu, Dong Yu and Pengyuan Liu 49016

UniGUI-Bench: A Real-World Diagnostic Benchmark — What Capabilities Do GUI Agents Most Need?
Jiaxing Liu, Ruoqi Sun, Liya Ma, FengJiao Chen, Ziwen Wang and Xuezhi Cao 49040

Dual-Uncertainty Guided Policy Learning for Multimodal Reasoning
Rui Liu, Dian Yu, Tong Zheng, Runpeng Dai, Zongxia Li, Wenhao Yu, Zhenwen Liang, Linfeng Song, Haitao Mi, Pratap Tokekar and Dong Yu 49054

How Do We Engage with Other Disciplines? A Framework to Study Meaningful Interdisciplinary Discourse in Scholarly Publications
Bagyasree Sudharsan, Alexandria Leto and Maria Leonor Pacheco 49070

CArtBench: A Museum-Grounded Benchmark for Evidence-Faithful Chinese Art Understanding
Xuefeng Wei, Wang Zhixuan, Xuan Zhou, Zhi Qu, Hongyao Li, Yusuke Sakai, Hidetaka Kamigaito and Taro Watanabe 49091

A Framework for Generating Valid Context-Specific Benchmarks through Expert Guidance
Kimberly Le Truong, Nari Johnson, Anna Kawakami and Hoda Heidari 49117

When AI Settles Down: Late-Stage Stability as a Signature of AI-Generated Text Detection
Ke Sun, Guangsheng Bao, Han Cui and Yue Zhang 49153

AgentCVR: Active Multi-Agent Cross-Video Reasoning via Script-Simulated Reinforcement Learning
Yilun Qiu, Jiahe Wang, Cilin Yan, Jiayin Cai, Xiaolong Jiang, Yao Hu and Chun Yuan 49180

Reformulating Radiology Report Error Detection as Sentence-Level Vision-Language Learning
Dabin Min, Kyungmin Jeon, Dongheon Lee and Chang Min Park 49212

ATTCK-R: Knowledge-Aware Long-Horizon Jailbreaking for Protected LLMs
Md Mushfiqur Rahman, Tobias Rohde, Rahul Tripathi, Khalil Elkhalil and Kevin Lybarger49229

Modeling Ranking Properties with In-Context Learning
Nilanjan Sinhababu, Andrew Parry, Debasis Ganguly and Pabitra Mitra 49256

StudyBench: Can Self-Evolution Squeeze Textbooks for Olympiad Capability?
Yinghao Chen, Zixi Chen, Bingxiang He, Ziqing Qiao, Huan-ang Gao, Yinuo Xu, Yuxin Zuo, Zeyuan Liu, Yuhao Zhan and Chaojun Xiao 49272

ARTKGC: Leakage-Controlled Multimodal Knowledge Graph Completion for Artwork Explanation
Xuan Zhou, Xuefeng Wei, Yuta Kato, Shintaro Ozaki, Kazuki Hayashi, Yusuke Sakai, Hidetaka Kamigaito, Katsuhiko Hayashi and Taro Watanabe 49291

CMAS: Enhancing Data Augmentation for Molecular Graph Classification via a Collaborative Multi-Agent System
Xuanting Fan, Chenyu Wang, Xin Liu, Weiyi Ren, Yifan Wang and Xiao Luo 49310

MulEG: Multi-Level Explanation-Guided Feature Attribute Graph for Zero-Shot Tabular Classification
Jiayang Zhang, Yunxiao Zhao, Xin Kang, Zhichao Yan, Boxiang Ma, Qinghua Chai, Xingtong Yu and Ru Li 49329

Towards Robust Retrieval-Augmented Generation via Reliability-Aware Adaptive Evidence Selection
Minkyu Che, Jihwan Oh, Jeonghwan Choi, Jeesu Jung and Hwanjun Song 49353

Hidden Thoughts Are Not Secret: Reasoning-Trace Exposure in LLMs
Yu-An Lu, Ci-Yang Tsai, Yu-Lin Tsai, Raluca Ada Popa and Chia-Mu Yu 49375

BhashaBench V1: A Comprehensive Benchmark for the Quadrant of Indic Domains
Vijay Devane, Mohd Nauman, Bhargav Patel, Aniket Mahendra Wakchoure, Yogeshkumar Sant, Shyam Pawar, Viraj Thakur, Ananya Godse, Sunil Patra, Neha Maurya, Suraj Racha, Nitish Kamal Singh, Piyush Sawarkar, Kundeshwar Vijayrao Pundalik, Rohit Saluja and Ganesh Ramakrishnan 49394

Hierarchical Wasserstein Merging for Multi-Domain Multi-Task Learning: From Specialists to a Generalist
Ming Cheng, Jiaying Gong and Hoda Eldardiry 49457

Anchored Reasoning on Tables: Bridging the Verifiability Gap in Multimodal Reinforcement Learning
Yubo Jiang, Yitong An, Fengying Xie, Zhiguo Jiang and Haopeng Zhang 49477

Lost in Embedding: Recovering Information Loss in Dense Retrieval
Prayas Agrawal and Niket Tandon 49498

Office Comprehension Benchmark
Firoz Shaik, Mateus Picanço Lima Gomes, Md Tanvir Islam Aumi, Jingci Wang, Milos Milunovic, Ivana Jovanovic, Jay Rathi, Vishwas Suryanarayanan, Neha Nandan Kenkare, Weiyao Xie, Zhipeng Han, Zheng Zhang, Waleed Shahid, Thong Q. Nguyen, Michael Bentley, Tamara Stankovic, Rasika Chakravarthy and Vishal Chowdhary 49518

Learning Pathology-Localized Medical Vision–Language Representations with Atomic and Grounded Disease Supervision
Mong Yuan Sim, Wei Emma Zhang, Xiang Dai, Byron Fang and Miao Liu 49550

RetroReasoner: A Reasoning LLM for Strategic Retrosynthesis Prediction
Hanbum Ko, Chanhui Lee, Ye Rin Kim, Rodrigo Hormazabal, Sehui Han, Sungbin Lim and Sungwoong Kim 49565

KARE-RAG: Knowledge-Aware Refinement and Enhancement for RAG
Yongjian Li, HaoCheng Chu, Yukun Yan, Zhenghao Liu, Shi Yu, Zheni Zeng, Ruobing Wang, Sen Song, Zhiyuan Liu and Maosong Sun 49601

Evaluating Stochastic Collapse and Implicit Bias in Multimodal Large Language Models
Huiyuan Zheng, Houtao Zhang, Boyang Wang, Qingyi Si and Hongcheng Guo 49624

SAGE: Satisfiability Guided Decoding for Mathematical Reasoning
Tamanna Hossain, Sameer Singh and Kareem Ahmed 49649

When More Metadata Misleads: Probing Field-Level Grounding in LLM-Generated Museum Descriptions
Clemencia Siro and Laura Hollink 49666

M^3 Scaling Law: Optimizing Multi-Epoch, Multi-Lingual, and Multi-Stage Training for Low-Resource Language Models
Kosuke Akimoto, Taiki Miyagawa and Masafumi Oyamada 49689

MedACE: A Time-Aware Asynchronous Clinical Environment for Long-Horizon Medical Treatment Decision-Making
Zhiting Fan, Ruizhe Chen, Hualiang Wang, Ru Peng, Zijie Meng, Wen MA, Jian Wu and Zuozhu Liu 49730

Lost in Reordering: Structural Sensitivity of Multilingual LLMs under Semantics-Preserving Perturbations
Karthika Nhayakkat, Rajat Verma, Maharaj Brahma, Vetcha Gnana Mahesh, Maunendra Sankar Desarkar, Ganesh Ramakrishnan and Rohit Saluja 49744

AdaptMMBench: Benchmarking Adaptive Multimodal Reasoning for Mode Selection and Reasoning Process
Xintong Zhang, Xiaowen Zhang, Jingrong Wu, Zhi Gao, Shilin Yan, Zhenxin Diao, Kunpeng Gao, Xuanyan Chen, Yuwei Wu, Yunde Jia and Qing Li 49761

Beyond Correlation-Driven Fusion: Sample-Adaptive Causal Evidence Learning for Multimodal Sentiment Analysis
Gaifang Luo, Hao Zhang, Yanan Liu, Haomin Tan, Guochuan Wu and Dan Xu 49799

Distribution-Consistent Inference for Dynamic Sparse Mixture-of-Experts
Dohyeon Kim, Bedionita Soro and Sung Ju Hwang 49812

MeDocVL: Noise-Aware Post-Training of Visual Language Models for Medical Document Parsing
Wei Wu, Liang Diao, Yuan Zhao, Wenjie Wang, Ying Liu and Xiaole Lv 49837

LoRi: Low-Rank Distillation for Implicit Reasoning
Ryan Solgi, Jiayi Tian and Zheng Zhang 49860

Atom-RAG: Atomic Fact-Centric Retrieval for Multi-Hop Reasoning
Seok Jun Hwang, Soohyeong Kim, JungHyoun Kim, Jeonghyeon Park and Yong Suk Choi 49872

Graph Topology Meets Energy-Driven Pathfinding: Preserving Contextual Consistency in Retrieval-Augmented Generation
Yufeng Li, Zaijing Li, Fuchen Zhang, Zongzong Wu, Ming Zhao and Fengxiao Tang 49891

SearchAtlas: Analyzing Agentic Search Strategies via Evidential Query Graphs
Jiacheng Sang, Mengyuan Li, Sanxing Chen, Yukun Huang, Yu Feng and Bhuwan Dhingra 49907

HERIT: Enabling Global Access to Hanja-to-English Historical Translation by Mitigating Temporal Bias
Dayeon Jeong and Kyeongpil Kang 49937

MVEB: Massive Video Embedding Benchmark
Adnan El Assadi, Roman Solomatin, Isaac Chung, Chenghao Xiao, Deep Shah, Shriya Sudhakar, Ayush Sunil Munot, Márton Kardos, Niklas Muennighoff and Kenneth Enevoldsen 49954

DEPART: DEcomposing PARiTy across Multilingual LLMs
Manan Uppadhyay, Prashant Kodali, Pranjal A Chitale, Reshma Ramaprasad, Himanshu Beniwal and Sunayana Sitaram 50003

To Copy or Not to Copy: Copying Is Easier to Induce Than Recall
Mehrdad Farahani, Franziska Penzkofer and Richard Johansson 50027

Beyond Global Scalars: Synergizing Token-Level Statistics and Deep Semantics for Adversarial AIGC Text Detection
Peiming Li, Yifan Wang, Zhiyuan Hu, Shiyu Li, Zheng Wei and Yang Tang 50091

Rollout-Level Advantage-Prioritized Experience Replay for GRPO
Gyeongtae Yoo, Sanghyeok Park, Soohyuk Jang, Ik-hwan Kim and Sungroh Yoon 50107

Proactive for Uncertainty: Cause-Aware Error Diagnosis and Interactive Clarification for Spoken Dialogue Systems
Yizhou Peng, Ziyang Ma, Changsong Liu, Yi-Wen Chao, Xie Chen and Eng Siong Chng . 50131

CliQ-RRG: Clinical-Knowledge Guided Disease-aware Visual–Textual Alignment for QA-Style Radiology Report Generation
Chandravardhan Singh Raghaw, Shahid Shafi Dar, Arnav Nishith Sharma, Shrey Modi and Nagendra Kumar 50153

An Honesty-Aware Multi-Agent Framework for Synthetic Psychiatric Intake Data Generation
Xinyuan Zhang, Chang Dao, Zijian Wang, Yechao Pang and Juexiao Zhou 50172

BuzzASR: A Swarm of 100+ Monolingual Speech Recognition Models
Shivam Singh, Aditya Yadavalli, Catherine Arnett and Alex Warstadt 50199

Writing As Reasoning: Interleaving Drafting and Deepening for Open-Ended Deep Research
Yishan Li, Wentong Chen, Yukun Yan, Mingwei Li, Kunpeng Liu, Xin Cong, Shuo Wang, Zhong Zhang, Yaxi Lu, Zhenghao Liu, Yankai Lin, Zhiyuan Liu and Maosong Sun 50228

An Empirical Analysis of Factual Errors in Human-Written Text and Its Application to Factual Error Detection
Kazuma Iwamoto, Kazumasa Omura and Shotaro Ishihara 50250

Rethinking the Multilingual Reasoning Gap with Layer Swap
Maxence Lasbordes, Amélie Chatelain and Djamé Seddah 50265

Router Prior Bias: Preserving Base Routing Structure in MoE Post-Training
Jaedeok Lee, Keonwoo Kim, Dongyoon Han, Sangdoo Yun, Yera Choi and Haanju Yoo . . 50281

Knowledge is Still There: Why RL Beats SFT for Sustainable Continual Machine Translation
Junhong Wu, Yang Zhao, Yangyifan Xu and Chengqing Zong 50302

ArenaFlow: From Trajectory Ranking to Hierarchical Credit Propagation for Open-Ended Agent RL
Qiang Zhang, Ruixue Ding, Fanrui Zhang, Xi Chen, Boli Chen, Shihang Wang, Yinfeng Huang, Yi Zheng, Pengjun Xie, Kaipeng Zhang, Jiawei Liu and Zheng-Jun Zha 50321

FigmaBench: Evaluating Design-to-Code Generation in Real-World Handoff Scenarios
Ziyang Wang, Ziyang Liu, Xin Yin, Chao Zhang, Zhe Cui and Yue Lu 50344

Same Evidence, Different Decisions: Document Segmentation Bias in Retrieval-Augmented Generation
Jiang Hui, Dichun Deng, Chenyang Zhou, Zhenrui Zhang and Chenghui Wu 50364

EvoAgentBench: Benchmarking Agent Self-Evolution via Ability Transfer
Xingze Gao, Chuanrui Hu, Hongda Chen, Pengfei Yao, Zhao Wang, Yi Bai, Zhengwei Wu, Yunyun Han, Xiaofeng Cong, Jie Gui, Yafeng Deng and Teng Li 50387

Teach a Man to Fish: Self-Adaptive Reinforced Distillation for In-Context Reasoning-based Reranker
Wenwei Wu, Youhang LI, Mingyang Chen, Jie Yang, Xiaofeng Gu, Duyi Pan, Yiwen Wu, Kaichen Zhang, Wei Wang and Zixuan Yuan 50403

The Art of Hierarchical Competing Patterns: Gaussian Process Optimization of Hyphenation
Ondřej Sojka and Petr Sojka 50440

Vishing-Tactics-Bench: Forecasting Exploitation Trajectories in Voice Phishing Calls
Jeongmin Lee, DongMyung Sul, Seung Yun and Jin-Xia Huang 50458

Counterfactual Tests for Measuring Chain-of-Thought Faithfulness in Visual Language Models
Bayar Menzat, Maximilian Süss, Ruizhi Wang, Benno Steinegger, Thomas Lukasiewicz and Oana-Maria Camburu 50474

Meeseeks-Multilingual: A Multi-Turn, Multi-Constraint Benchmark evaluating Multilingual Instruction-Following and Self-Correction
Jiaming Wang, Hefei Chen, Zhe Tang, Zehao Jin, Yanxin Zhuo, Ruicheng Liao, Jinrong Ma, Xiaoyu Li and Xuezhi Cao 50505

LeakageBench: Document-Level Leakage Risk for Redacting Personally Identifiable Information in Document Images
Vishnu Prasad Vijaya Kumar, Santhosh Venkatesh and Ivan P. Yamshchikov 50519

HAPMoE: Heterogeneity-Aware Automatic Parallelism Planning for Mixture-of-Experts Models Training
Mengyuan Fan, Peizhuang Cong, Zixiao Huang, Si Xu, Tong Qiao, Yanghao Li, Jing Yang, Tong Yang, Quanlu Zhang and Yu Wang 50538

MoMem: A Mixture-of-Memory Architecture for Balancing Factual Recall and Experiential Reuse in LLM Agents
Shuo Sun, Ke Rui and Xuehai Hong 50553

Any-Order in Theory, Left-to-Right in Practice: The Autoregressive Underpinnings of dLLMs
Dayuan Zhao and Yiduo Guo 50571

CompEvo: Competition-Induced Evolution for Multi-Agent in News-Driven Time Series Forecasting
Yuxuan Zhang, Yangyang Feng, Yong Guan, Daifeng Li, Kexin Zhang, Junlan Chen, Bowen Deng, Jun Liu and Zehua Zeng 50581

Democratic ICAI: Debating Our Way to Steering Principles from Preferences
Kevin Kingslin, Anish Natekar, Ashutosh Ranjan, Vivek Srivastava, Savita Bhat and Shirish Karande 50608

Sieve and Sage: Efficient Distraction Filtering for Reliable RALM Abstention
Jongbin Won, SungGeun AN and Jay-Yoon Lee 50655

FinAcumen: Financial Multimodal Reasoning via Self-Evolving Experience Memory Harness
Pianran Guo, Pengcheng Zhou, Yucheng Jian, Shuhua Chen, Zhongliang Yang and Linna Zhou 50675

Encoders Strike Back: Why and How a Strong Discriminative Baseline Outperforms LLMs in Argument Mining Across Three Heterogeneous Datasets
Fatma-Zohra Rezkellah, Ghassen Karray and Reto Gubelmann 50696

Latent Reasoning Guidance for Parallel Code Translation
Tomer Bitan, Erel Kaplan, Roee Bar-Yadin, Lian Ghrayeb, Le Chen, Samyak Jhaveri, Niranjan Hasabnis and Gal Oren 50714

OARelatedWork: A Large-Scale Dataset of Related Work Sections with Full-texts from Open Access Sources
Martin Docekal, Martin Fajčík and Pavel Smrz 50728

LiveClawBench: Toward Faithful Evaluation of Stateful LLM Assistants
Xiang Long, Li Du, Yilong Xu, RongJian Xu, Qiyanhui LU, Ying Gao, Qinhua Xie, Fangcheng Liu, Ning Ding, Haoqing Wang, Ziheng Li, Changjiang Zhou, Jianyuan Guo and Yehui Tang . . . 50753

Computationally Invisible: Charting the NLP Landscape of Iranic Languages
Hojjat Mokhtarabadi, Amir Hossein Kargaran, Ali Shakeri, Saba Shafiei, Nafiseh Nikeghbal and Mohammad Hossein Manshaei 50767

Singing from the Same ANTHEM: An Enriched ACL Anthology for Scholarly Comparison
Moritz Staudinger, Wojciech Kusa, Pia Pachinger and Allan Hanbury 50798

BNPO: Beta Normalization Policy Optimization
Changyi Xiao, Mengdi Zhang and Yixin Cao 50817

LLM-Generated or Human-Written? Comparing Review and Non-Review Papers on ArXiv
Yanai Elazar and Maria Antoniak 50830

Linguistic Distance Segregates Latent Representations in Automatic Speech Recognition Systems
Ting-Hui Cheng, Line H Clemmensen and Sneha Das 50849

"Act Like a 5th Grader" is Not Enough: Bounding Knowledge in LLM-Based User Simulators
Krisztian Balog and Arild Michel Bakken 50860

CLIN: an Objective Framework for Evaluating Creativity in Short Persian Literary Text
Mohammad Reza Modarres, Armin Tourajmehr, Yadollah Yaghoobzadeh and Mohammad Taher Pilehvar 50879

Can Dialects Be Steered Like Languages? Sparse Neurons and Distributed Directions in Arabic LLMs
Kareem Elozeiri, Mervat Abassy, Omar Kallas, Fahim Dalvi, Preslav Nakov, Kentaro Inui and Nadir Durrani 50894

Probing Fine-Grained Stereotypes and Trait Structure in Large Language Models
Rebecca Yu, Nicholas Deas, Colin Wayne Leach and Kathleen McKeown 50919

SUPERNOVA: Eliciting General Reasoning in LLMs with Reinforcement Learning on Natural Instructions
Ashima Suvarna, Kendrick Phan, Mehrab Beikzadeh, Hritik Bansal and Saadia Gabriel . . . 50948

PICO Entity Datasets and Methodologies in the Context of Randomized Controlled Trials: A Case Study
Julien Bezançon, Pierre Zweigenbaum and Nona Naderi 50973

How Does Alignment Tuning Shape Representations of Sycophancy and Related Cue-Induced Biases in LLMs?
Prakhar Gupta, Terry Jingchen Zhang, Florent Draye, Bernhard Schölkopf and Zhijing Jin 50984

H^2-MRG: Hierarchical Heterogeneous Graph Based Medical Report Generation with Fine-Grained Disease-Specific Modeling
Libiao Chen, Dong Nie and Zhenyu Tang 51002

GraP-Mem: Granularity-Aware Planning for Adaptive Memory Access in Long-Horizon LLM Agents
Qinyang Zeng, Jinhao Dong, Wei LU and Xiaoyong Du 51016

FLOUR: Mixing the Right Ingredients for LLM Unlearning in Latent Space
Thuy Dung Nguyen, Giang Nguyen, Taylor T Johnson and Kevin Leach 51033

Metaphors are a Source of Cross-Domain Misalignment of Large Reasoning Models
Zhibo Hu, Chen Wang, Yanfeng Shu, Hye-young Paik and Liming Zhu 51052

Towards Natural Personalization: Evaluating Long-Horizon Preference Following in Personalized User-LLM Interactions
Qianyun Guo, Yibo Li, Yue Liu and Bryan Hooi 51070

State2State: Revealing Compositional Safety Failures through State-Transition Reasoning
Chu-Jun Peng, Zhong-Wei Lin, Yu-Wen Tseng, Hong-Han Shuai and Wen-Huang Cheng . 51106

What Do Developers Ask? A Taxonomy of Code Related Questions
Ashutosh Ojha, Ashita Saxena, Varshini Reddy Pateel, Palanivel Kodeswaran, Sayandeep Sen and Srikanth G. Tamilselvam 51122

Do We Still Need Humans in the Loop? Human vs. LLM Annotation in Active Learning for TikTok Hate Speech Detection
Ahmad Dawar Hakimi, Lea Hirlimann, Isabelle Augenstein and Hinrich Schuetze 51141

Can LLMs Model Incorrect Student Reasoning? A Case Study on Distractor Generation for Multiple-Choice Questions
Yanick Zengaffinen, Andreas Opedal, Donya Rooein, Kv Aditya Srivatsa, Shashank Sonkar and Mrinmaya Sachan 51171

GTA-RAG: Graph-Trajectory-Augmented Reinforcement Learning for Multi-Turn Retrieval-Augmented Reasoning
Jun Chen, Yongchao Liu, Pengyu Qiu, Jiajun Zheng, Juelu Zhang, Yujie Zeng, Qin Zhang, Ziyue Qiao and Xiao Luo 51207

FinHopper: A Progressive Multi-Hop Benchmark for Complex Financial Agents with Step-wise Evaluation
Yuxuan Wu, Miaopei Lin, Qiufang Ying, Lin Dan, Hao Zhao, Qirui Liu, Yuteng Shen, Shaohui Wu, Weihongluo, Xiku Du, Xing Tang and Hao Chen 51219

To Retrieve or To Think? Cross-Boundary Context Evolution for Multi-hop Complex Reasoning
Rubing Chen, Jian Wang, Wenjie Li, Xiao-Yong Wei and Li Qing 51247

What Limits Us? Analyzing Self-Reported Limitations in NLP Research
Tawan Thaepprasit, Peeranuth Kehasukcharoen, Ding Wang, Remi Denton, Peerapon Vateekul and Piyawat Lertvittayakumjorn 51260

Ask Better Questions: Learning Efficient Information-Seeking Policy for Clinical Cold-Start Dialogue
Li Wenbo, Le Cai, Jiawei Liu, Jinhong Yu, Zhuo Chen, Zhuoran Luo, Yang Qingqing, Gao Cheng, Qikai Cheng, Wei Lu and Xiaozhong Liu 51298

Are Multilingual Models Improving? Revisiting Cross-Lingual Transfer
Prasoon Bajpai, Eleftheria Briakou, Preethi Jyothi and Vihari Piratla 51329

Can Vibe Coders Really Pass the Vibe Check? An Empirical Analysis of Fault-Targeted Reasoning in Frontier LLMs
Srijan Bansal, Fangkai Jiao, Yilun Zhou, Austin Xu, Shafiq Joty and Semih Yavuz 51358

MSI-SQL: Unleashing Data Potential of Large Language Models for Text-to-SQL via Multi-Task Self-Iterative Generation
Li Gao, Zixuan Wang, Boci Peng, Mingjun Nie, Huan Luo and Chun Gan 51372

DSPP: Dataset-Specific Progressive Pruning for Efficient and Generalizable LLM Deployment
Lichuan Wan, Huiqiang Jiang, Wenxuan Liu and Hua Yang 51383

En-ViMedNER: An English-Vietnamese Parallel Biomedical Corpus with UMLS Semantic Type Annotations
Nhu Vo, Phuong Nguyen, Nu Uyen Phuong Le, Inigo Jauregi Unanue, Dung D. Le, Massimo Piccardi and Wray Buntine 51397

On Mitigation of Subliminal Learning
Atsushi Yanagisawa, Brendan Gho, Rajendran Ramesh Babu Manoj Narender, Joel Jacob, Kevin Zhu, Madhur Panwar and Antonio Mari 51419

Reassessing High-Performing LLMs on Polish Medical Exams: True Competence or Bias-Driven Performance?
Antoni Lasik, Jakub Pokrywka, Łukasz Grzybowski, Jeremi Ignacy Kaczmarek, Gabriela Korzańska, Janusz Świeczkowski-Feiz, Oskar Pastuszek, Paulina Hoffman, Jakub Tomasz Dabrowski and Wojciech Kusa 51441

VICT: Verifier-Instrumented Credit Tracing for Reinforcement Learning of Long-Horizon LLM Agents
Pengcheng Li, Zhengyang Zhang, Dongxu Zhang, Sui Huang and Shaohua Ma 51468

CircuitLens: Reasoning Circuits as Data Selection Signals for Reinforcement Learning with Verifiable Rewards
Zhuofan Chen, Ziqian Jiao, Yikai Cui, Zhixin Cai, Jun Bai and Wenge Rong 51487

Is Visual Prompting All You Need? Studying VLM Spatial Reasoning under Progressive Visual Scaffolds
Lars Benedikt Kaesberg, Tianyu Yang, Florian Valentin Wunderlich, Terry Ruas, Daniel Kurzawe, Jan Philip Wahle and Bela Gipp 51505

Interleaved Speech Language Models Latently Work In Text
Talia Sternberg, Gallil Maimon and Yossi Adi 51527

Provenance-Grounded Gating and Adaptive Recovery in Synthetic Post-Training Data Curation
Soham Bhattacharjee, Karun Sharma, Vinay Kumar Sankarapu and Pratinav Seth 51556

My own morality, my own mind": Persona-Conditioned Moral and Cultural Reception on Trump Messages
Candida Maria Greco, Víctor Yeste, Andrea Tagarelli and Paolo Rosso 51567

Discriminative World Models for Web Agents
Kelvin Li, Dhruv Pendharkar, Anish Pahilajani, Chuyi Shang, Leon Oks, Leonid Karlinsky, Rogerio Feris, Trevor Darrell and Roei Herzig 51596

ReBound: Relevance-Boundary Supervision for Distractor-Robust Mathematical Reasoning
Jiahui Long, Lidan Shou, Huan Li and Ke Chen 51612

What Matters When Building Universal Multilingual Named Entity Recognition Models?
Jonas Golde, Patrick Haller and Alan Akbik 51628

Beyond Prompts: A Systematic Study of LLM-Based Machine Translation Evaluators
Dmitry Popov, Roman Bokhyan, Ekaterina Enikeeva, Artem Mekhraliev, Stepan Vysotsky and Nikolay Karpachev 51648

Routing-Aware Head Amplification for Multimodal Long-Chain-of-Thought Reasoning
Chuhao Zhou, Zonghan Wu, Haolang Lu, Kun Wang, Weiran Yao, Linsey Pang, Prakhar Mehrotra, Qingsong Wen and Yilei Shao 51665

Higher-Dimensional Rotary Position Embedding
Yixing Li, Ruobing Xie, Yudong Zhang, Yushi Bai, Samm Sun and Yu Cheng 51677

Agent Memory Distillation: Empowering Small LLM Agents with Hierarchical Teacher Memory
Taeil Kim, Kangsan Kim and Sung Ju Hwang 51693

Relevance as a Vulnerability: How Web Retrieval Degrades Safety Alignment in LLM Agents
Aditya Nawal, Manit Baser and Mohan Gurusamy 51720

SimplexLoRA: Dynamic Rank Updating via Alternating Minimization with Simplex Projection
Shalygin Igor, Grigorii Davydenko, Andrey Veprikov, Aleksey Antonov, Nikolay Bushkov, Stanislav Moiseev and Aleksandr Beznosikov 51747

How Far Can Machine Translation Quality Take You? Extrinsic Discourse Evaluation in Goal-Oriented Setups
Wafaa Mohammed, Kata Naszadi and Vlad Niculae 51757

CARVE: Verified Expansion for Variable-Length Generation in Diffusion Language Models
Wail Bouhedja, Amr Mohamed and Guokan Shang 51779

LLMs Decide Early and Explain Later: You Can Stop Them Early
Ayan Datta, Zhixue Zhao, Mounika Marreddy, Bhuvanesh Verma, Radhika Mamidi and Alexander Mehler 51796

KnowVis: Knowledge-Centric Visual Summarization for Video Lectures
Yi Xu, Yifan Hou and Xiaoyu Zhang 51823

Understanding Failures in LLM Reasoning by Learning Structured Representations of Chain-of-Thought
Tommaso Felice Banfi, Mubashara Akhtar, Jingwei Ni and Mrinmaya Sachan 51841

Ladders in Chaos: When, How, (and Perhaps Why) Does Test-Time Scaling Improve LLM Machine Translation
Di Wu, Sergey Troshin, Christof Monz, Antske Fokkens and Vlad Niculae 51863

When Clinical Notes Are Incomplete: Fusing Clinical Notes and Time Series Using Diffusion-based Models
Feilong Wu, Kejing Yin and William K. Cheung 51894

Enhancing SAE-based Steering via Neighbor Integrated Feature Selection
Yutian Liu, Xu Wang and Difan Zou 51914

LLM Forensics: Where Do Backdoors Hide? Localizing and Controlling Trigger Mechanisms with Sparse Autoencoders
Wissam Antoun, Francis Kulumba, Théo Lasnier, Benoît Sagot and Djamé Seddah 51929

TrustMem: Learning Trustworthy Memory Consolidation for LLM Agents with Long-Term Memory
Tianyu Yang, Sudipta Paul, Vijay Srinivasan, Vivek Kulkarni and Srinivas Chappidi 51948

CALIBER: Calibrating Confidence Before and After Reasoning in Language Models
Conor Finlay, Joshua Kurien, Saurabh Dash, Marzieh Fadaee and Beyza Ermis 51966

CI-RELAY: Contextual Integrity Mediation for Privacy in LLM-based Multi-Agent Systems
Runshan Hu, Yukun Dong, Jia Bi and Chenyang Zhu 51981

Closing the Cross-Dialect Gap: Query Plans as a Portable Interface in Text-to-SQL
Corentin Royer, Robin Oester, Yotam Perlitz, Yannick Metz, Andrea Giovannini and Mennatallah El-Assady 52002

Investigating Query-Insensitive Behavior in Spatio-Temporal Video Grounding
Eryk Kołodziejczyk, Alberto Presta, Karol Szurkowski and Michal Byra 52022

From Language Specifications to Executable Turing Machines: Evaluating LLMs as Computational Machine Designers
Shahriyar Zaman Ridoy, S. M. Muhtasimul Hasan, Azmine Toushik Wasi, Kinga Tshering, Koushik Ahamed Tonmoy and Md Mosaddek Khan 52031

Informed Masking: Structure-Aware Perturbation for Reinforcement Learning in Diffusion Large Language Models
Xiaoyi Yu, Enver Sangineto, Pei Fu, Fiorenzo Parascandolo, Wenhui Tan, Ruikang Zhang, Rita Cucchiara, Ruihua Song and Jian Luan 52053

Investigating Assistant Bias in LLM User Simulators Using a Role Vector
Daeheon Jeong, Yoonjoo Lee, Eugene Choi, Sinie van der Ben and Juho Kim 52070

WEREWOLF: Reputation-Aware Red-Teaming for Self-Organizing LLM Multi-Agent Systems
Manhin Poon, Qirun Zeng, Xiangxiang Dai and Jinhang Zuo 52105

Targeting the Exceptions: Online Active Learning for Retrieval-Based LLM Routers
Clovis Varangot-Reille, Antoine Gourru, Christophe Bouvard and Baptiste Jeudy 52117

GMTS: Gradient Magnitude-based Token Selection Improves RLVR Training for LLM Reasoning
Outongyi Lv, Yuanwei Zhang and Xiaoqun Zhang 52136

Variable-Rank Adaptation: A Comprehensive Survey of Rank Budgeting LoRA Variants
Max Ploner and Alan Akbik 52150

Exponential-Tilted Unlearning
Sungwon Chae, Keonwoo Kim and Hoki Kim 52173

MoralTrajectories: Evaluating Moral Steerability in Large Language Models
Daniel Russo, Marco Antonio Stranisci, Oscar Araque and Marco Guerini 52200

Quantifying Retriever-Generator Alignment in RAG with Local Explanations
Korbinian Randl, Guido Rocchietti, Aron Henriksson, Ziawasch Abedjan, Tony Lindgren and John Pavlopoulos 52217

SMMU: Benchmarking Social Intelligence of Multimodal Large Language Models
Gordon Chen, Ziqi Huang, Ailing Zeng, Chuhan Mei, Kairui Hu, Xin Tong and Ziwei Liu 52242

Trust the Uncertain Teacher: Distilling Dark Knowledge via Calibrated Uncertainty
Jeonghyun Kim, Richeng Xuan, Hyung Wook Noh, Soo Kyung Kim and Hyunsoo Cho . . . 52255

How Far Can Current Tokenizers Go in Bangla? A Benchmark and a Morphology-Aware Solution
Gautam Siddharth Kashyap, MD Azizul Hoque, Kinjalk Srivastava, Pratik Jalan, Subhashis Kar and Usman Naseem 52275

Who Argues What? Joint Argument-Entity Detection and Classification in Political Debates
Lucio La Cava, Stefano Francesco Monea and Sergio Greco 52285

Text-Based Veracity Assessment of Verbal Statements in Swedish
Clara Gustafsson and Joakim Nivre 52303

Evolving Agents in the Dark: Retrospective Harness Optimization via Self-Preference
Wenbo Pan, Shujie Liu, Chin-Yew Lin, Jingying Zeng, Xianfeng Tang, Xiangyang Zhou, Yan Lu and Xiaohua Jia 52321

GUIDE: Guiding Internal Evidence with Language Instructions
Caren Han, Hyunsuk Chung, Jinwoo Kim, Seungyeon Ji and Kyungreem Han 52363

To MRL or not to MRL: Text Embeddings are Robust to Truncation Without Matryoshka Learning, Except In Heavy Truncation Scenarios
Sotaro Takeshita, Yurina Takeshita, Simone Paolo Ponzetto and Daniel Ruffinelli 52385

Do Neural Retrievers Prefer Certain Documents? Evidence of Learned Relevance Priors
Francisco Valentini, Edgar Altszyler and Martin Fajčík 52405

Cross-Lingual Exploration for Parametric Knowledge
Elisha Diskind, Itamar Trainin, Uri Shaham, Leshem Choshen, Idan Szpektor and Omri Abend 52429

End-to-End Differentiable Optimization of Tool-Augmented Visual Reasoning via Probabilistic Graphs
Wentao Wan, Kaiyu Wu, Qingyang Ma, Nan Kang, Yunjie Chen, Liang Lin and Keze Wang52467

Algospeak, Hiding in the Open: The Trade-off Between Legible Meaning and Detection Avoidance
Jan Fillies, Ronald E. Robertson and Jeffrey Hancock 52483

From AI Assistant to AI Scientist: Autonomous Discovery of LLM-RL Algorithms with LLM Agents
Sirui Xia, Yikai Zhang, Aili Chen, Siye Wu, Siyu Yuan and Yanghua Xiao 52498

Can LLMs Imagine Moral Alternatives Beyond Binary Dilemmas?
Jongchan Choi, Nari Yang, Sung Soo Park, Jaemin Cho, Han Seoyoung, Haerin Shin and Jun-Hyung Park 52524

Definition-Only RAG for Financial QA: Average Gains and a Diagnostic Boundary
Leyi Ouyang, Edison Marrese-Taylor and Yutaka Matsuo 52551

VFPO: Variance-aware Negative Feedback Policy Optimization for Mitigating Entropy Collapse via Group Statistics
Bo Zhang, Zinan Zeng, Juntong Zhou, Dading Chong, Zhenyu Ding, Dengpeng Wu, Ning Ding and Peilin Zhou 52561

Ican easily jailbreak LLMs
Eunkyung Choi, Yeeun Kim, Juheon Kang and Wonseok Hwang 52589

Knowing When to Speak: Aligning LLM Honesty via Solvability-Aware Preference Optimization
Wenpeng Lu, Xueguan Zhao, Shuxia Lin, Linhai Zhang, Chaoqun Zheng, Weiyu Zhang and Jiasheng Si 52609

Don't Start from Scratch: Memory-Informed Detoxification for Safer LLM Responses
Pengze Chen, Jiachuan Wang, Weijie Sun, Wangze Ni and Lei Chen 52627

Measure what Matters: Psychometric Evaluation of AI with Situational Judgment Tests
Alexandra Yost, Shreyans Jain, Shivam Raval, Grant C. Corser, Nina Xu, Allen G Roush, Jacqueline Hammack, Ravid Shwartz Ziv and Amir Abdullah 52645

Can We Stop Malicious AI? KILLBENCH: A Benchmark for External AI Kill Switch Feasibility
Sechan Lee, Hyounghun Kim and Sangdon Park 52751

From Answers to Interpretations: Rethinking Ambiguity-Induced Aleatoric Uncertainty Estimation in LLMs
Omer Nahum, Niv Nayman, Jonathan Fhima, Alon Zolfi, Jeremy Levy, Shai Mazor and Paolo Favaro 52799

When Patients Cut In: Extending Clinical Conversational AI Safety to Interruptions
Zachary Ellis, Spencer Hazel, Adam Brandt, Ernest Lim, Yajie Vera He and Jared Joselowitz 52828

IndicSafeEval: Safety Robustness of Large Language Models under Multilingual Persuasive Jailbreak Attacks
Saikat Mondal, Mamta Mamta, Deeksha Varshney, Oana Cocarascu and Asif Ekbal 52853

Bye Bye Perspective API: Lessons for Building and Governing Measurement Infrastructure
David Hartmann, Manuel Tonneau, Angelie Kraft, LK Seiling, Dimitri Staufer, Pieter Delobelle, Jan Fillies, Anna Ricarda Luther, Jan Batzner and Mareike Lisker 52890

GAPD: Gold-Action Policy Distillation for Agentic Reinforcement Learning in Knowledge Base Question Answering
Xin Sun, Jianan Xie, Zhongqi Chen, Qiang Liu, Shu Wu, Bowen Song, Weiqiang Wang, Zilei Wang and Liang Wang 52915

Probing Large Audio-Language Models for Compositional Understanding of Sounding Actions
Michel Olvera, Paraskevas Stamatiadis, Changhong Wang and Gaël Richard 52936

Beyond Turn Limits: Training Deep Search Agents with Dynamic Context Window
Qiaoyu Tang, Hao Xiang, Le Yu, Bowen Yu, Yaojie Lu, Xianpei Han, Le Sun, Wenjuan Zhang, Pengbo Wang, Shixuan Liu, Zhenru Zhang, Jianhong Tu, Hongyu Lin and Junyang Lin 52951

EBS: Efficient Block-Sparse Framework for Continual Learning of Large Language Models
Zhixin Han, Hang Gao, Yinhao Bai, Yuhua Zhao, Bitong Luo, Zhiyu li, Mengting Hu and Minlie Huang 52983

Internalizing Negation-Gated Logical Rules into LLMs for Document-Level Relation Extraction
Hye-Yoon Baek, Sangjun Ji, Jimyeung Seo, Hae-Yoon Koo, Xiongnan Jin and Byungkook Oh 52994

Some Assembly Required: Pointer-Decoded Seq2Seq Models for German Without a Native Seq2Seq Backbone
Jan Pfister, Tom Völker and Andreas Hotho 53012

What Tokens are Learned when Tokenization is Optimized Jointly with Language Modeling?
Saketh Reddy Vemula and Parameswari Krishnamurthy 53028

Replacing Training with Memory: Listwise Selection for Text-to-SQL
Yeonseok Jeong, Soyoung Yoon, Seongjun Lee and Seung-Won Hwang 53049

Multi-turn Conversational AI from Text to Multimodal Interaction: Data, Models, Evaluation, and Open Challenges
Syeda Faiza Ahmed Sara, Zien Sheikh Ali, Hunzalah Hassan Bhatti, Firoj Alam and Shammur Absar Chowdhury 53068

KCSAT-ML: Probing Reasoning Models with Nationwide-Cohort Human Difficulty
Sanghee Park, Geewook Kim and Kee-Eung Kim 53106

One Rate Is Not Enough: Adaptive Anisotropic Learning Rates for LoRA Fine-Tuning
Huiyi Wang, Daijiao Liu, Lina Yao and Dong Gong 53130

Omni-Embed-Mini: Binding Modalities Without Forgetting via Dense Distillation
Mohammed Irfan Kurpath, Jaseel Muhammad Kaithakkodan, Sahal Shaji Mullappilly, Ivan Laptev and Hisham Cholakkal 53150

EvoBrowseComp: Benchmarking Search Agents on Evolving Knowledge
Yunhan Wang, Jiaan Wang, Lianzhe Huang, Xianfeng Zeng and Fandong Meng 53176

Soft Sequence Policy Optimization
Svetlana Glazyrina, Maksim Kryzhanovskiy and Roman Ischenko 53191

ExCAM: Explainable Cultural Awareness Metrics
Christoph Leiter, Haiyue Song, Hour Kaing, Jin Tei, Hideki Tanaka, Masao Utiyama and Steffen Eger 53202

Why Latent Chain-of-Thought Helps Exploration but Hurts Precise Computation
Jiaxuan Zou, Yaozhong Xiong and Yong Liu 53231

PRISM-VLM: A Multi-Axis Discriminative Benchmark for Compact Vision-Language Models
Sanghee Park and Kee-Eung Kim 53259

CPF: Concept-based Probabilistic Fingerprinting for Copyright Protection in Multimodal Large Language Models
Xinghao Wang, Xiaofan Zheng, Peng Peng and Minnan Luo 53288

BAITBENCH: Measuring Agent Reward Hacking with Optional Shortcuts Planted in ML Tasks
Pradyumna Shyama Prasad, Meiri Anto, Leon Eshuijs, Julian Moncarz, Kaustubh Kislay and Juan J Vazquez 53305

A Survey on Entropy Mechanism in Large Reasoning Models
Zihan Wang, Xingle Xu, Hao Wang, Yiwen Ye, Yuchen Li, Linhao Wang, Hongze Tan, Peidong Wang, Shi Feng, Guoqing Chen, Jinghao Lin, Zijing Wang, Yiqun Zhang, Yongkang Liu, Xiaocui Yang, Tao Yan, Lei Yang, Shengzhi Wang, Yuhang Wu and Ge Yu 53321