

ACL 2025

**The 63rd Annual Meeting of the Association for  
Computational Linguistics (ACL 2025)**

**Proceedings of the Conference – Volume 6: Industry Track**

July 28-30, 2025

©2025 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

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

ISBN 979-8-89176-288-6

## Message from the ACL 2025 Industry Track Co-Chairs

We are happy and excited to welcome you to the Industry Track at ACL 2025, held on the three main days of the 63rd Annual Meeting of the Association for Computational Linguistics (28 July to 30 July 2025).

Like the main research track, the industry track attracted an unprecedented number of submissions: 421 papers! In total, 453 reviewers and 19 area chairs participated in the evaluation of these papers. After a thorough, double-blind peer-review evaluation with three reviews for each submission followed by reviewer discussions and additional deliberations, 108 papers were selected for presentation at the ACL 2025 Industry Track. Of these, 33 papers will be presented as oral talks and a total of 75 papers will be presented as posters.

Topic-wise, large language models were front and center of almost all submissions with trustworthiness, domain-adaptation, retrieval-augmented generation, and agentic architectures – across domains such as medical, legal, and finance – being popular topics.

NLP research in academia and NLP research in industry have always been very close in our fields. Our two keynote speakers – Lucia Specia and Leon Derczynski – will share their insights with regard to this intersection including synergies and fruitful collaborations.

Further insights can be gained through our “Careers in NLP” panel with esteemed participants who have decades of experience with NLP research in academia and industry.

We would like to thank the authors of all Industry Track submissions as well as the reviewers and area chairs for their hard and dedicated work under very tight deadlines. We would also like to thank the General Chair, the Publication Chairs, who supported us in the production of this volume, and all other ACL 2025 committees we interacted with between the summer of 2024, when this endeavour started, and the summer of 2025, when we finally have been able to have the Industry Track at the ACL 2025 conference in Vienna, Austria. Finally, we would also like to thank our keynote speakers and panellists as well as the whole ACL team, especially Jennifer Rachford.

Georg Rehm and Yunyao Li

Program Co-Chairs

# Program Committee

## Program Chairs

Georg Rehm, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI) and Humboldt-Universität zu Berlin, Germany  
Yunyao Li, Adobe, USA

## Area Chairs

Trung Bui, Adobe Research  
Marina Danilevsky, IBM Research  
Dejing Dou, Fudan University  
Xue-Yong Fu  
Jose Manuel Gomez-Perez, expert.ai  
Fatima Haouari, University of Sheffield  
Leonhard Hennig, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)  
Jussi Karlgren, AMD Silo AI  
Varun Kumar, Amazon  
Amita Misra, Amazon  
Ani Nenkova, Adobe Research  
Cennet Oguz, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)  
Kun Qian, Adobe Systems  
Vera Schmitt, Technische Universität Berlin  
Björn Schuller, Technische Universität München and Imperial College London  
Andrejs Vasiljevs, Tilde and University of Latvia  
Chenguang Wang, Washington University, Saint Louis  
François Yvon, ISIR, Sorbonne Université & CNRS  
Honglei Zhuang, Google DeepMind

## Reviewers

Ahmed Abdelali, Asad Abdi, Raia Abu Ahmad, Sallam Abualhaija, Abdalghani Abujabal, Manoj Acharya, Amit Agarwal, Sachin Agarwal, Sanjay Agrawal, Mohammad Shahmeer Ahmad, Reem AlYami, Firoj Alam, Jan Alexandersson, Georgios Alexandridis, Kenneth Alperin, Duygu Altinok, Duygu Altinok, Somya Anand, Dimitra Anastasiou, Rafael Anchiêta, Enrique Henestroza Anguiano, Mario Ezra Aragon, Eiji Aramaki, Kushagr Arora, Ekaterina Artemova, Ankit Arun, Sören Auer

Long Bai, Mithun Balakrishna, Indervir Singh Banipal, Aditya Bansal, Daniel Bauer, Ian Beaver, Dorothee Beermann, Vibha Belavadi, Tadesse Destaw Belay, Cristian Berrio Aroca, Dario Bertero, Arjun Bhalla, Sumit Bhatia, Nikhil Bhendawade, Debmalya Biswas, Shivangi Bithel, Emmanuel Aboah Boateng, Ekaterina Borisova, Nadjet Bouayad-Agha, Chris Brew, Thomas Brovelli, Yi Bu, Paul Buitelaar

Lucas Pereira Carlini, Fabio Casati, Thiago Castro Ferreira, Yekun Chai, Raghuveer Chanda, Ciprian Chelba, Cheng Chen, Fuxiang Chen, Guanhua Chen, Jiangning Chen, Jiangning Chen, John Chen, Lei Chen, Lihu Chen, Lin Chen, Luoxin Chen, Qingyu Chen, Wei Chen, Yiru Chen, Zhi-Qi Cheng, Shamil Chollampatt, Jaegul Choo, Yun-Wei Chu, Ann Clifton, Simone Conia, Rylan Conway, Georgiana Copil, Bonaventura Coppola, Giandomenico Cornacchia

Deborah A. Dahl, Ayushi Dalmia, Arion Das, Souvik Das, Tirthankar Dasgupta, Anderson De Andrade, Steve DeNeefe, Lingjia Deng, Shumin Deng, Xingjian Diao, Daniel Dickinson, Bo-sheng Ding, Rahul Divekar, Bin Dong, Li Dong, Zi-Yi Dou, Eduard Dragut, Andrew Drozdov, Matthew T. Dunn, Sourav Dutta

Jessica Maria Echterhoff, Aparna Elangovan, Heba Elfardy, David Elson

Michael Flor, Simona Frenda, Lisheng Fu, Yanjie Fu, Gilad Fuchs

Ankur Gandhe, Pengyu Gao, Radhika Gaonkar, Carmen Garcia-Mateo, Federico Gaspari, Xiou Ge, Anna Lisa Gentile, Ryan Georgi, Piyush Ghai, Diman Ghazi, Sayan Ghosh, Voula Giouli, Anmol Goel, Ethan Goh, Jiaying Gong, Nidhi Goyal, Annika Grützner-Zahn, Kalpa Gunaratna, Beliz Gunel, Honglei Guo, Tong Guo, Ankush Gupta, Raghav Gupta, Amit Gupte, Ramiro H. Gálvez

Amir Hadifar, Udo Hahn, Benjamin Han, Songyang Han, Youssef Al Hariri, Ulrich Heid, Amr Hendy, Sanjika Hewavitharana, Derrick Higgins, Swapnil Hingmire, Mengze Hong, Pengyu Hong, Kristen Howell, Guimin Hu, Seung-won Hwang

Akshay Jagatap, Utkarsh Jain, Miloš Jakubiček, Ala Jararweh, Janet Jenq, Aastha Jhunjunwala, Meng Jiang, Ishan Jindal, Rosie Jones

Hiroshi Kanayama, Jun Seok Kang, Yashal Shakti Kanungo, Raghav Kapoor, Pinar Karagoz, Yanis Katsis, Yoav Katz, Hamit Kavas, Roman Kern, Elena Khasanova, Byung-Hak Kim, Geewook Kim, Tracy Holloway King, Miyoung Ko, Thomas H Kober, Svetla Peneva Koeva, Sai Koneru, Fajri Koto, George Kour, Jared Kramer, Rajasekar Krishnamurthy, Soundarya Krishnan, Marek Kubis, Sanjeev Kumar, Sanjeev Kumar, Vinayshekhar Bannihatti Kumar, Tzu-Lin Kuo, Igor Kuzmin

Yanis Labrak, Sarasi Lalithsena, Stefan Larson, Md Tahmid Rahman Laskar, Md Tahmid Rahman Laskar, Alexandra Lavrentovich, Daniel Lee, Jaeseong Lee, Young-Suk Lee, Deren Lei, Jochen L. Leidner, Elena Leitner, Brian Lester, Ran Levy, Changmao Li, Dingcheng Li, Dongfang Li, Mingda Li, Pengyuan Li, Qiang Li, Xupeng Li, Yuyang Li, Gilbert Lim, Nut Limsopatham, Antonie Lin, Ting-En Lin, Ying Lin, Guangliang Liu, Hanchao Liu, Jingyuan Liu, Jingyuan Liu, Siyang Liu, Xuye Liu, Ye Liu, Mengsay Loem, Jaime Lorenzo-Trueba, Natalia V Loukachevitch, Jiaming Luo, Wencan Luo, Ziyang Luo, Teresa Lynn, Alexander Lyzhov

Mingyuan MA, Nianzu Ma, Mounica Maddela, Akash V Maharaj, Wolfgang Maier, Fred Mailhot, Piotr Mardziel, Eugenio Martínez-Cámara, Santosh Mashetty, Santosh Mashetty, Puneet Mathur, Yuji Matsumoto, Evgeny Matusov, David D. McDonald, Alexander Mehler, Kartik Mehta, Helen M. Meng, Fabio Mercorio, Md Messal Monem Miah, Margot Mieskes, Nandana Mihindukulasooriya, Hideya Mino, Katya Mirylenka, Hemant Misra, Abubakr Mohamed, Shekoofeh Mokhtari, Julian Moreno Schneider, Yasmin Moslem, Sidharth Mudgal, Abhishek Mukherji, Matthew Mulholland, Syed Shariyar Murtaza, John Murzaku, Emir Muñoz

Masaaki Nagata, Manish Nagireddy, Tetsuji Nakagawa, Jinseok Nam, Tarek Naous, Diane Napolitano, Krishnasuri Narayanam, Yaroslav Nechaev, Dmitry Nikolaev, Nobal B. Niraula, Navid Nobani, Elnaz Nouri

Alexander O'Connor, Oleg Okun, Eda Okur, Sergio Oramas, Naoki Otani

Ankur Padia, Alonso Palomino, Dookun Park, Haeju Park, Jeiyoon Park, Jeiyoon Park, Seong-Jin Park, Youngja Park, Ioannis Partalas, Sangameshwar Patil, Priyaranjan Pattnayak, Bibek Paudel, Stephan Peitz, Thang M. Pham, Mārcis Pinnis, Jakub Piskorski, Saloni Potdar, Arantza Del Pozo, Vishnu Prabhakaran, Shishir Kumar Prasad, Radityo Eko Prasajo, Gábor Prószéky, Michal Ptaszynski, Stephen Pulman

Long Qin, Xin Ying Qiu, Elio Querze

Taki Hasan Rafi, Sajjadur Rahman, Nitin Ramrakhiyani, Leonardo Ranaldi, Prasanjit Rath, Venkatesh Ravichandran, Meghana Ravikumar, Vipula Rawte, Traian Rebedea, Georg Rehm, Ehud Reiter, Maarten de Rijke, Matiss Rikters, Joe Cheri Ross, Sumegh Roychowdhury, Weitong Ruan, Mukund Rungta

Harald Sack, Alicia Sagae, Tanay Kumar Saha, Rishav Sahay, Avinash Sahu, Prathusha Kameswara Sarma, Sheikh Muhammad Sarwar, Felix Sasaki, Minoru Sasaki, Minoru Sasaki, Nayan Saxena, Kevin Scaria, David Schlangen, Andreas Schwarz, Frank Seide, Ethan Selfridge, Husrev Taha Sencar, Shubhashis Sengupta, Ramtin M. Seraj, Muhammad Shakeel, Yuan Shangguan, Sanat Sharma, Serge Sharoff, Qiang Sheng, Ashish Shenoy, Kejian Shi, Michal Shmueli-Scheuer, Ingo Siegert, Patrick Simianer, Sneha Singhania, Priyanka Sinha, Inguna Skadina, Kazoo Sone, Hyun-Je Song, Yang Song, Yuanfeng Song, Makesh Narsimhan Sreedhar, Arvind Krishna Sridhar, Manisha Srivastava, Sebastian Steindl, Evgeny Stepanov, Sebastian Stüker, Chenkai Sun, Weixuan Sun, Marek Suppa, Shiv Surya, Sandesh Swamy, Jäder Martins Camboim De Sá

Santosh T.y.s.s, Yuki Tagawa, Sudarshan R. Thitte, Philippe Thomas, Wee Hyong Tok, Manabu Torii, Giuliano Tortoreto, Giuliano Tortoreto, Keith Trnka, Masaaki Tsuchida, Siddharth Tumre

Arun Palghat Udayashankar, Emmanuel Ngue Um, David Uthus

Praneetha Vaddamanu, Tom Vanallemeersch, Andrea Varga, Prasoon Varshney, Cristina Vertan, Ngoc Phuoc An Vo, Ngoc Phuoc An Vo, Piek Vossen, Duy-Khanh Vu

Bingqing Wang, Hai Wang, Jin Wang, Jun Wang, Kevin Shukang Wang, Mingxian Wang, Ryan Wang, Sitong Wang, Suge Wang, Yi-Chia Wang, Yile Wang, Yu Wang, Yun-Cheng Wang, Zhengxiang Wang, Penghui Wei, Kaixin Wu, Tianxing Wu

Kaige Xie, Zejun Xie, Hongyan Xu

Victor Yang, Dezhi Ye, Rong Ye, Jinyeong Yim, Yuwei Yin, Lei Yu

Fadi Zaraket, Qingkai Zeng, Dan Zhang, Kai Zhang, Kang Zhang, Lei Zhang, Tianlin Zhang, Yifan Zhang, Yin Zhang, Zhe Zhang, Zhehao Zhang, Fuheng Zhao, Dong Zhou, Zhengyu Zhou, Gao yu Zhu, Jennifer Zhu, Jiahao Zhu, Su Zhu, Wei Zhu, Xiliang Zhu, Bowei Zou

# Keynote Talk

## From Words to Worlds: NLP for Game Creation and Interaction

**Lucia Specia**

Epic Games and Imperial College London

**Mon, July 28th, 2025 – Time: 11:00 – 11:45 – Room: Austria Center Vienna**

**Abstract:** The gaming industry is a leading force in global entertainment, surpassing the size of the music and film industries combined. With over 3 billion people playing games, there's bigger and bigger demand for fresh, high-quality, and immersive experiences. At the same time, user-generated games have become a core component of major gaming platforms, fostering creativity and diversification. These developments present significant opportunities for AI research and AI-driven tools designed to support gaming, from AAA studios to independent creators. In this talk, I will highlight some of these opportunities, focusing on three areas involving language: 1) Speech-driven animation, where we predict lip sync, expression, and head motion of a character from audio to animate photo-realistic characters; 2) Low-resource language code generation, where we build a code generation model for Verse, a new language designed specifically for programming interactive 3D worlds, games, and simulations; and 3) Safety of interactive NPCs at scale, where we design safety strategies to support the deployment of LLMs for speech to speech in-game (Fortnite) conversations between players and NPCs.

**Bio:** Lucia Specia is Senior Director of Research Engineering at Epic Games and Professor of Natural Language Processing at Imperial College London. Her work focuses on various aspects of data-driven approaches to multimodal and multilingual context models, with applications including machine translation, image captioning, visual question answering, quality estimation, and content moderation, among others. In 2021, she founded Contex.ai to build multimodal content moderation models for real world applications, focusing in the gaming industry. She now leads a team of research engineers at Epic Games delivering ML solutions across automation and business optimization, safety and security, user experience and content creation. She received a PhD from the University of Sao Paulo and has held positions at University of Sheffield, Meta, and Xerox Research.

# Keynote Talk

## We can't do it alone

**Leon Derczynski**

NVIDIA and IT University of Copenhagen

**Tue, July 29th, 2025 – Time: 10:30 – 11:15 – Room: Austria Center Vienna**

**Abstract:** Industry and academic research each have their own deficiencies and blindspots, and both rely heavily on each other. This talk explores common themes and describes each side's view and what they miss for each theme. We will discuss the role in society, the role in research, which narratives work (and don't), the role in peer review (and its role for us), and where the hard workers, sceptics, and sociopaths fit in either case. All this comes together to form a positive view of good open collaborations, and some concrete advice on how to give and get the most value out of interactions with the other side.

**Bio:** Leon Derczynski is principal research scientist for LLM security at NVIDIA and prof in computer science at ITU Copenhagen. He has written inches, if not kilograms, of papers, and won similar quantities of awards etc. Prof. Derczynski has led policy efforts in academia, industry, and civil society. He has held affiliations at a dozen organisations in the past decade, including startups, universities, corporations, and non-profits; built research programmes at both a leading university and a leading corporation; and he retains a deep love of both university and industry research.



## Table of Contents

<i>ACL 2025 Industry Track: Overview</i>	
Georg Rehm and Yunyao Li .....	1
<i>Speculative Reward Model Boosts Decision Making Ability of LLMs Cost-Effectively</i>	
Jiawei Gu and Shangsong Liang .....	4
<i>RAVEN: Robust Advertisement Video Violation Temporal Grounding via Reinforcement Reasoning</i>	
Deyi Ji, Yuekui Yang, Haiyang Wu, Shaoping Ma, Tianrun Chen and Lanyun Zhu .....	22
<i>DistilQwen2.5: Industrial Practices of Training Distilled Open Lightweight Language Models</i>	
Chengyu Wang, Junbing Yan, Yuanhao Yue and Jun Huang .....	32
<i>SimUSER: Simulating User Behavior with Large Language Models for Recommender System Evaluation</i>	
Nicolas Bougie and Narimawa Watanabe .....	43
<i>Scaling Context, Not Parameters: Training a Compact 7B Language Model for Efficient Long-Context Processing</i>	
Chen Wu and Yin Song .....	61
<i>MathAgent: Leveraging a Mixture-of-Math-Agent Framework for Real-World Multimodal Mathematical Error Detection</i>	
Yibo Yan, Shen Wang, Jiahao Huo, Philip S. Yu, Xuming Hu and Qingsong Wen .....	69
<i>Towards Multi-System Log Anomaly Detection</i>	
Boyang Wang, Runqiang Zang, Hongcheng Guo, Shun Zhang, Shaosheng Cao, Donglin Di and Zhoujun Li .....	83
<i>LLM-Enhanced Self-Evolving Reinforcement Learning for Multi-Step E-Commerce Payment Fraud Risk Detection</i>	
Bo Qu, Zhurong Wang, Daisuke Yagi, Zach Xu, Yang Zhao, Yinan Shan and Frank Zahradnik .....	92
<i>ORMind: A Cognitive-Inspired End-to-End Reasoning Framework for Operations Research</i>	
Zhiyuan Wang, Bokui Chen, Yinya Huang, Qingxing Cao, Ming He, Jianping Fan and Xiaodan Liang .....	104
<i>Multi-Step Generation of Test Specifications using Large Language Models for System-Level Requirements</i>	
Dragan Milchevski, Gordon Frank, Anna Hättö, Bingqing Wang, Xiaowei Zhou and Zhe Feng .....	132
<i>RUBRIC-MQM : Span-Level LLM-as-judge in Machine Translation For High-End Models</i>	
Ahrii Kim .....	147
<i>SocialForge: simulating the social internet to provide realistic training against influence operations</i>	
Ulysse Oliveri, Guillaume Gadek, Alexandre Dey, Benjamin Costé, Damien Lolive, Arnaud Delhay and Bruno Grilheres .....	166
<i>TN-Eval: Rubric and Evaluation Protocols for Measuring the Quality of Behavioral Therapy Notes</i>	
Raj Sanjay Shah, Lei Xu, Qianchu Liu, Jon Burnsky, Andrew Bertagnolli and Chaitanya Shivade .....	179
<i>Run LoRA Run: Faster and Lighter LoRA Implementations</i>	
Daria Cherniuk, Aleksandr Mikhalev and Ivan Oseledets .....	200

<i>Genetic Instruct: Scaling up Synthetic Generation of Coding Instructions for Large Language Models</i>	
Somshubra Majumdar, Vahid Noroozi, Mehrzad Samadi, Sean Narenthiran, Aleksander Ficek, Wasi Uddin Ahmad, Jocelyn Huang, Jagadeesh Balam and Boris Ginsburg . . . . .	208
<i>NeKo: Cross-Modality Post-Recognition Error Correction with Tasks-Guided Mixture-of-Experts Language Model</i>	
Yen-Ting Lin, Zhehuai Chen, Piotr Zelasko, Zhen Wan, Xuesong Yang, Zih-Ching Chen, Krishna C Puvvada, Ke Hu, Szu-Wei Fu, Jun Wei Chiu, Jagadeesh Balam, Boris Ginsburg, Yu-Chiang Frank Wang and Chao-Han Huck Yang . . . . .	222
<i>Generating OpenAPI Specifications from Online API Documentation with Large Language Models</i>	
Koren Lazar, Matan Vetzler, Kiran Kate, Jason Tsay, David Boaz, Himanshu Gupta, Avraham Shinnar, Rohith D Vallam, David Amid, Esther Goldbraich, Jim Laredo and Ateret Anaby Tavor . .	237
<i>CoAlign: Uncertainty Calibration of LLM for Geospatial Repartition</i>	
Zejun Xie, Zhiqing Hong, Wenjun Lyu, Haotian Wang, Guang Wang and Desheng Zhang . . .	254
<i>Arctic-TILT: Business Document Understanding at Sub-Billion Scale</i>	
Łukasz Borchmann, Michał Pietruszka, Wojciech Jaśkowski, Dawid Jurkiewicz, Piotr Halama, Paweł Józiać, Łukasz Garncarek, Paweł Liskowski, Karolina Szyndler, Andrzej Gretkowski, Julita Ołtusek, Gabriela Nowakowska, Artur Zawłocki, Łukasz Duhr, Paweł Dyda and Michał Turski . . .	264
<i>Graph-Linguistic Fusion: Using Language Models for Wikidata Vandalism Detection</i>	
Mykola Trokhymovych, Lydia Pintscher, Ricardo Baeza-Yates and Diego Sáez Trumper . . .	284
<i>LOTUS: A Leaderboard for Detailed Image Captioning from Quality to Societal Bias and User Preferences</i>	
Yusuke Hirota, Boyi Li, Ryo Hachiuma, Yueh-Hua Wu, Boris Ivanovic, Marco Pavone, Yejin Choi, Yu-Chiang Frank Wang, Yuta Nakashima and Chao-Han Huck Yang . . . . .	295
<i>CiteFix: Enhancing RAG Accuracy Through Post-Processing Citation Correction</i>	
Harsh Maheshwari, Srikanth Tenneti and Alwarappan Nakkiran . . . . .	310
<i>Light-RL: Curriculum SFT, DPO and RL for Long COT from Scratch and Beyond</i>	
Liang Wen, Yunke Cai, Fenrui Xiao, Xin He, Qi An, Zhenyu Duan, Yimin Du, Junchen Liu, Tanglifu Tanglifu, Xiaowei Lv, Haosheng Zou, Yongchao Deng, Shousheng Jia and Xiangzheng Zhang	318
<i>Efficient Out-of-Scope Detection in Dialogue Systems via Uncertainty-Driven LLM Routing</i>	
Álvaro Zaera, Diana Nicoleta Popa, Ivan Sekulic and Paolo Rosso . . . . .	328
<i>Transforming Podcast Preview Generation: From Expert Models to LLM-Based Systems</i>	
Winstead Zhu, Ann Clifton, Azin Ghazimatin, Edgar Tanaka and Ward Ronan . . . . .	336
<i>A Perspective on LLM Data Generation with Few-shot Examples: from Intent to Kubernetes Manifest</i>	
Antonino Angi, Liubov Nedoshivina, Alessio Sacco, Stefano Braghin and Mark Purcell . . . .	345
<i>TablePilot: Recommending Human-Preferred Tabular Data Analysis with Large Language Models</i>	
Deyin Yi, Yihao Liu, Lang Cao, Mengyu Zhou, Haoyu Dong, Shi Han and Dongmei Zhang .	355
<i>LogicQA: Logical Anomaly Detection with Vision Language Model Generated Questions</i>	
Yejin Kwon, Daeun Moon, Youngje Oh and Hyunsoo Yoon . . . . .	411
<i>Model Merging for Knowledge Editing</i>	
Zichuan Fu, Xian Wu, Guojing Li, Yingying Zhang, Yefeng Zheng, Tianshi Ming, Yejing Wang, Wanyu Wang and Xiangyu Zhao . . . . .	433

<i>HierGR: Hierarchical Semantic Representation Enhancement for Generative Retrieval in Food Delivery Search</i>	
Fuwei Zhang, Xiaoyu Liu, Xinyu Jia, Yingfei Zhang, Zenghua Xia, Fei Jiang, Fuzhen Zhuang, Wei Lin and Zhao Zhang . . . . .	444
<i>Overlapping Context with Variable-Length Stride Increases Diversity when Training Large Language Model for Code</i>	
Geonmo Gu, Jaeho Kwak, Haksoo Moon, Hyun Seung Shim, Yu Jin Kim, Byoungjip Kim, Moon-tae Lee and Hyejeong Jeon . . . . .	456
<i>Generating Q&amp;A Benchmarks for RAG Evaluation in Enterprise Settings</i>	
Simone Filice, Guy Horowitz, David Carmel, Zohar Karnin, Liane Lewin-Eytan and Yoelle Maarek . . . . .	469
<i>Grammar-Constrained Decoding Makes Large Language Models Better Logical Parsers</i>	
Federico Raspanti, Tanir Ozcelebi and Mike Holenderski . . . . .	485
<i>AUTOSUMM: A Comprehensive Framework for LLM-Based Conversation Summarization</i>	
Abhinav Gupta, Devendra Singh, Greig A Cowan, N Kadhiresan, Siddharth Srivastava, Yagneswaran Sriraja and Yoages Kumar Mantri . . . . .	500
<i>RedactOR: An LLM-Powered Framework for Automatic Clinical Data De-Identification</i>	
Praphul Singh, Charlotte Dzialo, Jangwon Kim, Sumana Srivatsa, Irfan Bulu, Sri Gadde and Krishnaram Kenthapadi . . . . .	510
<i>Conceptual Diagnostics for Knowledge Graphs and Large Language Models</i>	
Rosario Uceda Sosa, Maria Chang, Karthikeyan Natesan Ramamurthy and Moninder Singh .	531
<i>QUPID: Quantified Understanding for Enhanced Performance, Insights, and Decisions in Korean Search Engines</i>	
Ohjoon Kwon, Changsu Lee, Jihye Back, Lim Sun Suk, Inho Kang and Donghyeon Jeon . . .	541
<i>Rethinking the Roles of Large Language Models in Chinese Grammatical Error Correction</i>	
Yinghui Li, Shang Qin, Jingheng Ye, Haojing Huang, Yangning Li, Shu-Yu Guo, Libo Qin, Xuming Hu, Wenhao Jiang, Hai-Tao Zheng and Philip S. Yu . . . . .	553
<i>EdgeInfinite: A Memory-Efficient Infinite-Context Transformer for Edge Devices</i>	
Jiyu Chen, Shuang Peng, Daxiong Luo, Fan Yang, Renshou Wu, Fangyuan Li and Xiaoxin Chen	568
<i>To Chat or Task: a Multi-turn Dialogue Generation Framework for Task-Oriented Dialogue Systems</i>	
Daniel Rim, Minsoo Cho, Changwoo Chun and Jaegul Choo . . . . .	576
<i>Scaling Under-Resourced TTS: A Data-Optimized Framework with Advanced Acoustic Modeling for Thai</i>	
Yizhong Geng, Jizhuo Xu, Zeyu Liang, Jinghan Yang, Xiaoyi Shi and Xiaoyu Shen . . . . .	593
<i>ArchiDocGen: Multi-Agent Framework for Expository Document Generation in the Architectural Industry</i>	
Junjie Jiang, Haodong Wu, Yongqi Zhang, Songyue Guo, Bingcen Liu, Caleb Chen Cao, Ruizhe Shao, Chao Guan, Peng Xu and Lei Chen . . . . .	605
<i>Optimization before Evaluation: Evaluation with Unoptimized Prompts Can be Misleading</i>	
Nicholas Sadjoli, Tim Siefken, Atin Ghosh, Yifan Mai and Daniel Dahlmeier . . . . .	619

<i>Think Again! The Effect of Test-Time Compute on Preferences, Opinions, and Beliefs of Large Language Models</i>	
George Kour, Itay Nakash, Michal Shmueli-Scheuer and Ateret Anaby Tavor . . . . .	639
<i>Learning from Litigation: Graphs for Retrieval and Reasoning in eDiscovery</i>	
Sounak Lahiri, Sumit Pai, Tim Weninger and Sanmitra Bhattacharya . . . . .	661
<i>LexGenie: Automated Generation of Structured Reports for European Court of Human Rights Case Law</i>	
Santosh T.y.s.s, Mahmoud Aly, Oana Ichim and Matthias Grabmair . . . . .	672
<i>Speed Without Sacrifice: Fine-Tuning Language Models with Medusa and Knowledge Distillation in Travel Applications</i>	
Daniel Zagyva, Emmanouil Stergiadis, Laurens Van Der Maas, Aleksandra Dokic, Eran Fainman, Ilya Gusev and Moran Beladev . . . . .	684
<i>Accelerating Antibiotic Discovery with Large Language Models and Knowledge Graphs</i>	
Maxime Delmas, Magdalena Wysocka, Danilo Gusicuma and Andre Freitas . . . . .	693
<i>Proactive Guidance of Multi-Turn Conversation in Industrial Search</i>	
Xiaoyu Li, Xiao Li, Li Gao, Yiding Liu, Xiaoyang Wang, Shuaiqiang Wang, Junfeng Wang and Dawei Yin . . . . .	706
<i>SpeechWeave: Diverse Multilingual Synthetic Text &amp; Audio Data Generation Pipeline for Training Text to Speech Models</i>	
Karan Dua, Puneet Mittal, Ranjeet Gupta and Hitesh Laxmichand Patel . . . . .	718
<i>Privacy Preserving Data Selection for Bias Mitigation in Speech Models</i>	
Alkis Koudounas, Eliana Pastor, Vittorio Mazzia, Manuel Giollo, Thomas Gueudre, Elisa Reale, Luca Cagliero, Sandro Cumani, Luca De Alfaro, Elena Baralis and Daniele Amberti . . . . .	738
<i>ComRAG: Retrieval-Augmented Generation with Dynamic Vector Stores for Real-time Community Question Answering in Industry</i>	
Qinwen Chen, Wenbiao Tao, Zhiwei Zhu, Mingfan Xi, Liangzhong Guo, Yuan Wang, Wei Wang and Yunshi Lan . . . . .	749
<i>PlanGPT: Enhancing Urban Planning with a Tailored Agent Framework</i>	
He Zhu, Guanhua Chen and Wenjia Zhang . . . . .	764
<i>FoodTaxo: Generating Food Taxonomies with Large Language Models</i>	
Pascal Wullschleger, Majid Zarharan, Donnacha Daly, Marc Pouly and Jennifer Foster . . . . .	784
<i>Enriching children's stories with LLMs: Delivering multilingual data enrichment for children's books at scale and across markets</i>	
Zarah Weiss, Christof Meyer and Mikael Andersson . . . . .	804
<i>Advanced Messaging Platform (AMP): Pipeline for Automated Enterprise Email Processing</i>	
Simerjot Kaur, Charese Smiley, Keshav Ramani, Elena Kochkina, Mathieu Sibue, Samuel Mensah, Pietro Totis, Cecilia Tilli, Toyin Aguda, Daniel Borrajo and Manuela Veloso . . . . .	813
<i>Semantic Outlier Removal with Embedding Models and LLMs</i>	
Eren Akbiyik, João F. M. De Almeida, Rik Melis, Ritu Sriram, Viviana Petrescu and Vilhjálmur Vilhjálmsón . . . . .	826
<i>SLENDER: Structured Outputs for SLM-based NER in Low-Resource Englishes</i>	
Nicole Ren and James Teo . . . . .	836

<i>A Large-Scale Real-World Evaluation of an LLM-Based Virtual Teaching Assistant</i> Sunjun Kweon, Sooyohn Nam, Hyunseung Lim, Hwajung Hong and Edward Choi . . . . .	850
<i>Operational Advice for Dense and Sparse Retrievers: HNSW, Flat, or Inverted Indexes?</i> Jimmy Lin . . . . .	865
<i>Filter-And-Refine: A MLLM Based Cascade System for Industrial-Scale Video Content Moderation</i> Zixuan Wang, Jinghao Shi, Hanzhong Liang, Xiang Shen, Vera Wen, Zhiqian Chen, Yifan Wu, Zhixin Zhang and Hongyu Xiong . . . . .	873
<i>ASK: Aspects and Retrieval based Hybrid Clarification in Task Oriented Dialogue Systems</i> Rishav Sahay, Lavanya Sita Tekumalla, Purav Aggarwal, Arihant Jain and Anoop Saladi . . . . .	881
<i>LEAP &amp; LEAN: Look-ahead Planning and Agile Navigation for LLM Agents</i> Nikhil Verma and Manasa Bharadwaj . . . . .	896
<i>MotiR: Motivation-aware Retrieval for Long-Tail Recommendation</i> Kaichen Zhao, Mingming Li, Haiquan Zhao, Kuien Liu, Zhixu Li and Xueying Li . . . . .	934
<i>A Framework for Flexible Extraction of Clinical Event Contextual Properties from Electronic Health Records</i> Shubham Agarwal, Thomas Searle, Mart Ratas, Anthony Shek, James Teo and Richard Dobson . . . . .	946
<i>Enhancing LLM-as-a-Judge through Active-Sampling-based Prompt Optimization</i> Cheng Zhen, Ervine Zheng, Jilong Kuang and Geoffrey Jay Tso . . . . .	960
<i>Small Language Models in the Real World: Insights from Industrial Text Classification</i> Lujun LI, Lama Sleem, Niccolo' Gentile, Geoffrey Nichil and Radu State . . . . .	971
<i>AutoChunker: Structured Text Chunking and its Evaluation</i> Arihant Jain, Purav Aggarwal and Anoop Saladi . . . . .	983
<i>User Feedback Alignment for LLM-powered Exploration in Large-scale Recommendation Systems</i> Jianling Wang, Yifan Liu, Yinghao Sun, Xuejian Ma, Yueqi Wang, He Ma, Zhengyang Su, Min- min Chen, Mingyan Gao, Onkar Dalal, Ed H. Chi, Lichan Hong, Ningren Han and Haokai Lu . . . . .	996
<i>SQLGenie: A Practical LLM based System for Reliable and Efficient SQL Generation</i> Pushpendu Ghosh, Aryan Jain and Promod Yenigalla . . . . .	1004
<i>Hard Negative Mining for Domain-Specific Retrieval in Enterprise Systems</i> Hansa Meghwani, Amit Agarwal, Priyaranjan Pattnayak, Hitesh Laxmichand Patel and Srikant Panda . . . . .	1013
<i>Interpretable Company Similarity with Sparse Autoencoders</i> Marco Molinari, Victor Shao, Luca Imeneo, Mateusz Mikolajczak, Abhimanyu Pandey, Vladimir Tregubiak and Sebastião Kuznetsov Ryder Torres Pereira . . . . .	1027
<i>Domain Adaptation of Foundation LLMs for e-Commerce</i> Christian Herold, Michael Kozielski, Tala Bazazo, Pavel Petrushkov, Yannick Versley, Seyyed Hadi Hashemi, Patrycja Cieplicka, Dominika Basaj and Shahram Khadivi . . . . .	1039
<i>sudo rm -rf agentic_security</i> Sejin Lee, Jian Kim, Haon Park, Ashkan Yousefpour, Sangyoon Yu and Min Song . . . . .	1050

<i>MedPlan: A Two-Stage RAG-Based System for Personalized Medical Plan Generation</i>	
Hsin-Ling Hsu, Cong-Tinh Dao, Luning Wang, Zitao Shuai, Thao Nguyen Minh Phan, Jun-En Ding, Chun-Chieh Liao, Pengfei Hu, Xiaoxue Han, Chih-Ho Hsu, Dongsheng Luo, Wen-Chih Peng, Feng Liu, Fang-Ming Hung and Chenwei Wu . . . . .	1072
<i>AIDE: Attribute-Guided Multi-Hop Data Expansion for Data Scarcity in Task-Specific Fine-tuning</i>	
Jiayu Li, Jennifer Zhu, Fang Liu and Yanjun Qi . . . . .	1083
<i>Synthesizing and Adapting Error Correction Data for Mobile Large Language Model Applications</i>	
Yanxiang Zhang, Zheng Xu, Shanshan Wu, Yuanbo Zhang and Daniel Ramage . . . . .	1102
<i>MultiMed: Multilingual Medical Speech Recognition via Attention Encoder Decoder</i>	
Khai Le-Duc, Phuc Phan, Tan-Hanh Pham, Bach Phan Tat, Minh-Huong Ngo, Thanh Nguyen-Tang and Truong-Son Hy . . . . .	1113
<i>MICE: Mixture of Image Captioning Experts Augmented e-Commerce Product Attribute Value Extraction</i>	
Jiaying Gong, Hongda Shen and Janet Jenq . . . . .	1151
<i>FINKRX: Establishing Best Practices for Korean Financial NLP</i>	
Guijin Son, Hyunwoo Ko, Hanearl Jung and Chami Hwang . . . . .	1161
<i>Sentiment Reasoning for Healthcare</i>	
Khai-Nguyen Nguyen, Khai Le-Duc, Bach Phan Tat, Le Duy, Long Vo-Dang and Truong-Son Hy	1175
<i>Judging the Judges: Can Large Vision-Language Models Fairly Evaluate Chart Comprehension and Reasoning?</i>	
Md Tahmid Rahman Laskar, Mohammed Saidul Islam, Ridwan Mahbub, Ahmed Masry, Mizanur Rahman, Amran Bhuiyan, Mir Tafseer Nayeem, Shafiq Joty, Enamul Hoque and Jimmy Huang . .	1203
<i>OccuTriage: An AI Agent Orchestration Framework for Occupational Health Triage Prediction</i>	
Alok Kumar Sahu, Yi Sun, Eamonn Swanton, Farshid Amirabdollahian and Abi Wren . . . .	1217
<i>One Missing Piece for Open-Source Reasoning Models: A Dataset to Mitigate Cold-Starting Short CoT LLMs in RL</i>	
Hyungjoo Chae, Dongjin Kang, Jihyuk Kim, Beong-woo Kwak, Sunghyun Park, Haeju Park, Jinyoung Yeo, Moontae Lee and Kyungjae Lee . . . . .	1227
<i>SingaKids: A Multilingual Multimodal Dialogic Tutor for Language Learning</i>	
Zhengyuan Liu, Geyu Lin, Hui Li Tan, Huayun Zhang, Yanfeng Lu, Xiaoxue Gao, Stella Xin Yin, Sun He, Hock Huan Goh, Lung Hsiang Wong and Nancy F. Chen . . . . .	1244
<i>Unifying Streaming and Non-streaming Zipformer-based ASR</i>	
Bidisha Sharma, Karthik Pandia D S, Shankar Venkatesan, Jeena J Prakash, Shashi Kumar, Malolan Chetlur and Andreas Stolcke . . . . .	1254
<i>A Semi-supervised Scalable Unified Framework for E-commerce Query Classification</i>	
Chunyu Yuan, Chong Zhang, Zhen Fang, Ming Pang, Xue Jiang, Changping Peng, Zhangang Lin and Ching Law . . . . .	1263
<i>CodeIF: Benchmarking the Instruction-Following Capabilities of Large Language Models for Code Generation</i>	
Kaiwen Yan, Hongcheng Guo, Xuanqing Shi, Shaosheng Cao, Donglin Di and Zhoujun Li .	1272
<i>BI-Bench : A Comprehensive Benchmark Dataset and Unsupervised Evaluation for BI Systems</i>	
Ankush Gupta, Aniya Aggarwal, Shivangi Bithel and Arvind Agarwal . . . . .	1287



<i>Reinforcement Learning for Adversarial Query Generation to Enhance Relevance in Cold-Start Product Search</i>	
Akshay Jagatap, Neeraj Anand, Sonali Singh and Prakash Mandayam Comar . . . . .	1300
<i>Auto Review: Second Stage Error Detection for Highly Accurate Information Extraction from Phone Conversations</i>	
Ayesha Qamar, Arushi Raghuvanshi, Conal Sathi and Youngseo Son . . . . .	1308
<i>From Recall to Creation: Generating Follow-Up Questions Using Bloom’s Taxonomy and Grice’s Maxims</i>	
Archana Yadav, Harshvivek Kashid, Medchalimi Sruthi, B JayaPrakash, Chintalapalli Raja Kulayappa, Mandala Jagadeesh Reddy and Pushpak Bhattacharyya . . . . .	1322
<i>A Parallelized Framework for Simulating Large-Scale LLM Agents with Realistic Environments and Interactions</i>	
Jun Zhang, Yuwei Yan, Junbo Yan, Zhiheng Zheng, Jinghua Piao, Depeng Jin and Yong Li .	1339
<i>ENGINius: A Bilingual LLM Optimized for Plant Construction Engineering</i>	
Wooseong Lee, Minseo Kim, Taeil Hur, Gyeong Hwan Jang, Woncheol Lee, Maro Na and Taeuk Kim . . . . .	1350
<i>Consistency-Aware Online Multi-Objective Alignment for Related Search Query Generation</i>	
Shuxian Bi, Chongming Gao, Wenjie Wang, Yueqi Mou, Chenxu Wang, Tang Biao, Peng Yan and Fuli Feng . . . . .	1365
<i>Towards Generating Controllable and Solvable Geometry Problem by Leveraging Symbolic Deduction Engine</i>	
Zhuoxuan Jiang, Tianyang Zhang, Peiyan Peng, Jing Chen, Yinong Xun, Haotian Zhang, Lichi Li, Yong Li and Shaohua Zhang . . . . .	1378
<i>TableCoder: Table Extraction from Text via Reliable Code Generation</i>	
Haoyu Dong, Yue Hu, Huailiang Peng and Yanan Cao . . . . .	1399
<i>Are LLMs reliable? An exploration of the reliability of large language models in clinical note generation</i>	
Kristine Ann M. Carandang, Jasper Meynard Arana, Ethan Robert Casin, Christopher Monterola, Daniel Stanley Tan, Jesus Felix B. Valenzuela and Christian Alis . . . . .	1413
<i>REVISE: A Framework for Revising OCRed text in Practical Information Systems with Data Contamination Strategy</i>	
Gyuho Shim, Seongtae Hong and Heuiseok Lim . . . . .	1423
<i>TaDA: Training-free recipe for Decoding with Adaptive KV Cache Compression and Mean-centering</i>	
Vinay Joshi, Pratik Prabhanjan Brahma, Zicheng Liu and Emad Barsoum . . . . .	1435
<i>Convert Language Model into a Value-based Strategic Planner</i>	
Xiaoyu Wang, Yue Zhao, Qingqing Gu, Zhonglin Jiang, Yong Chen and Luo Ji . . . . .	1444
<i>MIRA: Empowering One-Touch AI Services on Smartphones with MLLM-based Instruction Recommendation</i>	
Zhipeng Bian, Jieming Zhu, Xuyang Xie, Quanyu Dai, Zhou Zhao and Zhenhua Dong . . . .	1457
<i>ConCodeEval: Evaluating Large Language Models for Code Constraints in Domain-Specific Languages</i>	
Mehant Kammakomati, Sameer Pimparkhede, Srikanth G. Tamilselvam, Prince Kumar and Pushpak Bhattacharyya . . . . .	1466

<i>Unveiling Dual Quality in Product Reviews: An NLP-Based Approach</i>	
Rafał Poświata, Marcin Michał Mironczuk, Sławomir Dadas, Małgorzata Grębowiec and Michał Perełkiewicz.....	1480
<i>Enhancing Marker Scoring Accuracy through Ordinal Confidence Modelling in Educational Assessments</i>	
Abhirup Chakravarty, Mark Brenchley, Trevor Breakspear, Ian Lewin and Yan Huang .....	1498
<i>A Practical Approach for Building Production-Grade Conversational Agents with Workflow Graphs</i>	
Chiwan Park, Wonjun Jang, Daeryong Kim, Aelim Ahn, Kichang Yang, Woosung Hwang, Jihyeon Roh, Hyerin Park, Hyosun Wang, Min Seok Kim and Jihoon Kang.....	1508
<i>EXPLAIN: Enhancing Retrieval-Augmented Generation with Entity Summary</i>	
Yaozhen Liang, Xiao Liu, Jiajun Yu, Zhouhua Fang, Qunsheng Zou, Linghan Zheng, Yong Li, Zhiwei Liu and Haishuai Wang.....	1520
<i>EcoDoc: A Cost-Efficient Multimodal Document Processing System for Enterprises Using LLMs</i>	
Ravi K. Rajendran, Biplob Debnath, Murugan Sankaradass and Srimat Chakradhar.....	1530