

COLING 2025

**The 31st International Conference on
Computational Linguistics**

Proceedings of the Industry Track

January 19 - 24, 2025

The COLING 2025 organizers gratefully acknowledge the support from the following sponsors.



Host Institution



Diamond



Winter School Sponsor



Platinum



Gold



Silver



Supporter



©2025 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 979-8-89176-197-1

Message from the Program Chairs

Welcome to the COLING 2025 Industry Track!

The COLING 2025 Industry Track provides a valuable venue for industry practitioners to showcase their real-world NLP applications and offers a bridge between academic and industrial research. The accepted papers cover a wide range of topics including:

- **Large Language Models (LLMs):** applications, evaluation, efficient training and finetuning, transfer learning, explainability and transparency, data augmentation and generation, and domain adaptation.
- **Natural Language Processing (NLP) applications:** such as entity recognition, question answering, and dialogue systems.
- **Multimodal Understanding:** visual understanding and question answering, multimodal generation, and visual reasoning.
- **Chatbots and Dialogue Systems:** customer support, voice assistants, intent understanding, and slot filling.
- **Edge Computing and Edge AI:** LLM on device, data labeling, and personalized classifiers.
- **Knowledge Representation and Management:** knowledge graph construction, domain-specific language modeling, and multimodal knowledge representation.

The quality and breadth of the Industry Track papers are noteworthy. In all, we received 166 submissions, which were reviewed by 126 reviewers from various industrial and academic institutions. Of the submitted papers, 69 papers were accepted for publication (acceptance rate of: 41.6%).

We have been honored to serve as the chairs of the Industry Track, and we extend our gratitude to Owen Rambow and Leo Wanner, the COLING 2025 general chairs, for their confidence and support.

We hope that you will find the papers insightful and inspiring.

Enjoy the conference!

Kareem Darwish, Apoorv Agarwal

Program Committee

Program Chairs

Kareem Darwish, aiXplain, USA

Apoorv Agarwal

Reviewers

AJ Stent, Abdelrahman Mohamed, Abubakr Mohamed, Ahmed Abdelali, Ahmed Ali, Alicia Sagae, Amit Gupte, Anastassia Loukina, Ankit Arun, Anmol Goel, Anuj Goyal, Ashish Shenoy, Avneesh Saluja, Ayushi Dalmia, Bowei Zou, Budhaditya Deb, Chandresh Maurya, Chanjun Park, Daniel Bauer, Dara Curran, David Elson, David Thulke, David Uthus, Deborah Dahl, Derrick Higgins, Dongmin Kim, Ehsaneddin Asgari, Ehud Reiter, Enrique Henestroza Anguiano, Evgeny Matusov, Fabio Mercorio, Feifei Pan, Firoj Alam, Francisco Garcia, Geewook Kim, Gilad Fuchs, Giuliano Tortoreto, Giuseppe Di Fabrizio, Go Inoue, Hamdy Mubarak, Hamit Kavas, Harsha Aduri, Haryo Akbarianto Wibowo, Hemant Misra, Hodong Lee, Igor Kuzmin, Ioannis Partalas, Jaime Lorenzo-Trueba, Jared Kramer, Jiangning Chen, Jingyuan Liu, Jinseok Nam, Jinyeong Yim, John Chen, John Murzaku, Jose Garrido Ramas, Kaicheng Wu, Kamer Yuksel, Kareem Darwish, Keith Trnka, Li Dong, Ling Tsou, Long Qin, Lorenzo Malandri, Majd Hawasly, Marek Suppa, Margot Mieskes, Matthew Mulholland, Matthew Spencer, Meghana Ravikumar, Michael Flor, Mohamed Al-Badrashiny, Mohammad Ahmad, Nadir Durrani, Navid Nobani, Ngoc Phuoc An Vo, Nidhi Goyal, Nikhil Rasiwasia, Nikoletta Basiou, Nitin Ramrakhiani, Octavian Popescu, Omid Ghahroodi, Owen Rambow, Pablo Duboue, Piyush Ghai, Prithiviraj Damodaran, Roman Van Der Krogt, Sachin Pawar, Sagnik Ray Choudhury, Sanat Sharma, Sandesh Swamy, Sanjeev Kumar, Sanjeev Kumar, Sanjika Hewavitharana, Sanyog Yadav, Sayantan Mitra, Sergio Oramas, Shishir Kumar Prasad, Shiv Surya, Shuangyong Song, Shubhashis Sengupta, Simone D'Amico, Somya Anand, Srinivas Bangalore, Stephen Pulman, Swapnil Hingmire, Thiago Castro Ferreira, Tirthankar Dasgupta, Tong Guo, Utkarsh Jain, Varun Nagaraj Rao, Viktória Ondrejová, Vinayak Karande, Wei Zhu, Weitong Ruan, Weiwei Guo, Xiao Liu, Xiliang Zhu, Xuye Liu, Yashal Shakti Kanungo, Yifan Zhang, Youssef Al Hariri, Yuji Matsumoto, Yuval Marton, Zhengxiang Wang, ilknur Durgar Elkahout

Table of Contents

<i>STAND-Guard: A Small Task-Adaptive Content Moderation Model</i> Minjia Wang, Pingping Lin, Siqi Cai, Shengnan An, Shengjie Ma, Zeqi Lin, Congrui Huang and Bixiong Xu	1
<i>Query-LIFE: Query-aware Language Image Fusion Embedding for E-Commerce Relevance</i> Hai Zhu, Yuankai Guo, Ronggang Dou and Kai Liu	21
<i>Improving Tool Retrieval by Leveraging Large Language Models for Query Generation</i> Mohammad Kachuee, Sarthak Ahuja, Vaibhav Kumar, Puyang Xu and Xiaohu Liu	29
<i>Know Your RAG: Dataset Taxonomy and Generation Strategies for Evaluating RAG Systems</i> Rafael Teixeira de Lima, Shubham Gupta, Cesar Berrospi Ramis, Lokesh Mishra, Michele Dolfi, Peter Staar and Panagiotis Vagenas	39
<i>RED-CT: A Systems Design Methodology for Using LLM-labeled Data to Train and Deploy Edge Linguistic Classifiers</i> David Farr, Nico Manzonelli, Iain Cruickshank and Jevin West	58
<i>Beyond Visual Understanding Introducing PARROT-360V for Vision Language Model Benchmarking</i> Harsha Vardhan Khurdula, Basem Rizk and Indus Khaitan	68
<i>PDC & DM-SFT: A Road for LLM SQL Bug-Fix Enhancing</i> Yiwen Duan, Yonghong Yu, Xiaoming Zhao, Yichang Wu and Wenbo Liu	76
<i>Multilingual Continual Learning using Attention Distillation</i> Sanjay Agrawal, Deep Nayak and Vivek Varadarajan Sembium	91
<i>FS-DAG: Few Shot Domain Adapting Graph Networks for Visually Rich Document Understanding</i> Amit Agarwal, Srikant Panda and Kulbhushan Pachauri	100
<i>OKG: On-the-Fly Keyword Generation in Sponsored Search Advertising</i> Zhao Wang, Briti Gangopadhyay, Mengjie Zhao and Shingo Takamatsu	115
<i>Best Practices for Distilling Large Language Models into BERT for Web Search Ranking</i> Dezhi Ye, Junwei Hu, Jiabin Fan, Bowen Tian, Jie Liu, Haijin Liang and Jin Ma	128
<i>Rationale-Guided Distillation for E-Commerce Relevance Classification: Bridging Large Language Models and Lightweight Cross-Encoders</i> Sanjay Agrawal, Faizan Ahemad and Vivek Varadarajan Sembium	136
<i>Automated Clinical Data Extraction with Knowledge Conditioned LLMs</i> Diya Li, Asim Kadav, Aijing Gao, Rui Li and Richard Bourgon	149
<i>Can Large Language Models Serve as Effective Classifiers for Hierarchical Multi-Label Classification of Scientific Documents at Industrial Scale?</i> Seyed Amin Tabatabaei, Sarah Fancher, Michael Parsons and Arian Askari	163
<i>EDAR: A pipeline for Emotion and Dialogue Act Recognition</i> Elie Dina, Rania Ayachi Kibech and Miguel Couceiro	175

<i>No Size Fits All: The Perils and Pitfalls of Leveraging LLMs Vary with Company Size</i>	
Ashok Urlana, Charaka Vinayak Kumar, Bala Mallikarjunarao Garlapati, Ajeet Kumar Singh and Rahul Mishra	187
<i>Predicting Fine-tuned Performance on Larger Datasets Before Creating Them</i>	
Toshiki Kuramoto and Jun Suzuki	204
<i>A Recipe For Building a Compliant Real Estate Chatbot</i>	
Navid Madani, Anusha Bagalkotkar, Supriya Anand, Gabriel Arnson, Rohini K. Srihari and Kenneth Joseph	213
<i>Geo-Spatially Informed Models for Geocoding Unstructured Addresses</i>	
Uddeshya Singh, Devanapalli Ravi Shankar, Gowtham Bellala and Vikas Goel	236
<i>Resource-Efficient Anonymization of Textual Data via Knowledge Distillation from Large Language Models</i>	
Tobias Deußer, Max Hahnbüch, Tobias Uelwer, Cong Zhao, Christian Bauckhage and Rafet Sifa	243
<i>Fine-Tuning Medium-Scale LLMs for Joint Intent Classification and Slot Filling: A Data-Efficient and Cost-Effective Solution for SMEs</i>	
Maia Aguirre, Ariane Méndez, Arantza del Pozo, Maria Ines Torres and Manuel Torralbo	251
<i>Enhancing Large Language Models for Scientific Multimodal Summarization with Multimodal Output</i>	
Zusheng TAN, Xinyi Zhong, Jing-Yu Ji, Wei JIANG and Billy Chiu	263
<i>"Stupid robot, I want to speak to a human!" User Frustration Detection in Task-Oriented Dialog Systems</i>	
Mireia Hernandez Caralt, Ivan Sekulic, Filip Carevic, Nghia Khau, Diana Nicoleta Popa, Bruna Guedes, Victor Guimaraes, Zeyu Yang, Andre Manso, Meghana Reddy, Paolo Rosso and Roland Mathis	276
<i>LLM Evaluate: An Industry-Focused Evaluation Tool for Large Language Models</i>	
Harsh Saini, Md Tahmid Rahman Laskar, Cheng Chen, Elham Mohammadi and David Rossouw	286
<i>Enhancing Future Link Prediction in Quantum Computing Semantic Networks through LLM-Initiated Node Features</i>	
Gilchan Park, Paul Baity, Byung-Jun Yoon and Adolffy Hoisie	295
<i>Page Stream Segmentation with LLMs: Challenges and Applications in Insurance Document Automation</i>	
Hunter Heidenreich, Ratish Dalvi, Nikhil Verma and Yosheb Getachew	305
<i>Graph-Augmented Open-Domain Multi-Document Summarization</i>	
Xiaoping SHEN and Yekun Chai	318
<i>Improve Speech Translation Through Text Rewrite</i>	
Jing Wu, Shushu Wang, Kai Fan, Wei Luo, Minpeng Liao and Zhongqiang Huang	331
<i>CarMem: Enhancing Long-Term Memory in LLM Voice Assistants through Category-Bounding</i>	
Johannes Kirmayr, Lukas Stappen, Phillip Schneider, Florian Matthes and Elisabeth Andre ...	343
<i>XTR meets ColBERTv2: Adding ColBERTv2 Optimizations to XTR</i>	
Riyaz Ahmad Bhat and Jaydeep Sen	358

<i>sDPO: Don't Use Your Data All at Once</i>	
Dahyun Kim, Yungi Kim, Wonho Song, Hyeonwoo Kim, Yunsu Kim, SANGHOON KIM and Chanjun Park	366
<i>Contextual ASR Error Handling with LLMs Augmentation for Goal-Oriented Conversational AI</i>	
Yuya Asano, Sabit Hassan, Paras Sharma, Anthony B. Sicilia, Katherine Atwell, Diane Litman and Malihe Alikhani	374
<i>Federated Retrieval Augmented Generation for Multi-Product Question Answering</i>	
Parshin Shojaee, Sai Sree Harsha, Dan Luo, Akash Maharaj, Tong Yu and Yunyao Li	387
<i>Luna: A Lightweight Evaluation Model to Catch Language Model Hallucinations with High Accuracy and Low Cost</i>	
Masha Belyi, Robert Friel, Shuai Shao and Atindriyo Sanyal	398
<i>Seeing Beyond: Enhancing Visual Question Answering with Multi-Modal Retrieval</i>	
Boqi Chen, Anuj Khare, Gaurav Kumar, Arjun Akula and Pradyumna Narayana	410
<i>AutoProteinEngine: A Large Language Model Driven Agent Framework for Multimodal AutoML in Protein Engineering</i>	
Yungeng Liu, Zan Chen, Yuguang Wang and Yiqing Shen	422
<i>Building a Family of Data Augmentation Models for Low-cost LLM Fine-tuning on the Cloud</i>	
Chengyu Wang, Yuanhao Yue, jun huang and Peng Wang	431
<i>Where do LLMs Encode the Knowledge to Assess the Ambiguity?</i>	
Hancheol Park and Geonmin Kim	445
<i>On the effective transfer of knowledge from English to Hindi Wikipedia</i>	
Paramita Das, Amartya Roy, Ritabrata Chakraborty and Animesh Mukherjee	453
<i>BackMATH: Towards Backward Reasoning for Solving Math Problems Step by Step</i>	
Shaowei Zhang and Deyi Xiong	466
<i>Deploying Multi-task Online Server with Large Language Model</i>	
Yincen Qu, Hengyue Liu, Kun Wang, Xiangying Dai, Xiaoou Lu, Hui Zhou and Chao Ma	483
<i>LLM-Friendly Knowledge Representation for Customer Support</i>	
Hanchen Su, Wei Luo, Yashar Mehdad, Wei Han, Elaine Liu, Wayne Zhang, Mia Zhao and Joy Zhang	496
<i>Leveraging Multilingual Models for Robust Grammatical Error Correction Across Low-Resource Languages</i>	
Divesh Ramesh Kubal and Apurva Shrikant Nagvenkar	505
<i>A Simple yet Efficient Prompt Compression Method for Text Classification Data Annotation Using LLM</i>	
Yiran Xie, Debin Xiao, Ping Wang and Shuming Liu	511
<i>AMAN: Agent for Mentoring and Assisting Newbies in MMORPG</i>	
Jeehyun Lee, Seung-Moo Yang and Won Ik Cho	522
<i>KARRIEREWEGE: A large scale Career Path Prediction Dataset</i>	
Elena Senger, Yuri Campbell, Rob van der Goot and Barbara Plank	533

<i>Transforming Code Understanding: Clustering-Based Retrieval for Improved Summarization in Domain-Specific Languages</i>	
Baban Gain, Dibyanayan Bandyopadhyay, Samrat Mukherjee, Aryan Sahoo, Saswati Dana, Palanivel Kodeswaran, Sayandeep Sen, Asif Ekbal and Dinesh Garg	546
<i>Is my Meeting Summary Good? Estimating Quality with a Multi-LLM Evaluator</i>	
Frederic Thomas Kirstein, Terry Lima Ruas and Bela Gipp	561
<i>Learning to Rewrite Negation Queries in Product Search</i>	
Mengtian Guo, Mutasem Al-Darabsah, Choon Hui Teo, Jonathan May, Tarun Agarwal and Rahul Bhagat	575
<i>LAW: Legal Agentic Workflows for Custody and Fund Services Contracts</i>	
William Watson, Nicole Cho, Nishan Srishankar, Zhen Zeng, Lucas Cecchi, Daniel Scott, Suchetha Siddagangappa, Rachneet Kaur, Tucker Balch and Manuela Veloso	583
<i>UR2N: Unified Retriever and Reranker</i>	
Riyaz Ahmad Bhat, Jaydeep Sen, Rudra Murthy and Vignesh P	595
<i>An Automatic Method to Estimate Correctness of RAG</i>	
Chi Zhang, Vivek V. Datla, Aditya Shrivastava, Alfy Samuel, Zhiqi Huang, Anoop Kumar and Daben Liu	603
<i>DaCoM: Strategies to Construct Domain-specific Low-resource Language Machine Translation Dataset</i>	
Junghoon Kang, Keunjoo Tak, Joungsu Choi, Myunghyun Kim, Junyoung Jang and Youjin Kang	612
<i>ChartGemma: Visual Instruction-tuning for Chart Reasoning in the Wild</i>	
Ahmed Masry, Megh Thakkar, Aayush Bajaj, Aaryaman Kartha, Enamul Hoque and Shafiq Joty	625
<i>LoRA Soups: Merging LoRAs for Practical Skill Composition Tasks</i>	
Akshara Prabhakar, Yuanzhi Li, Karthik Narasimhan, Sham Kakade, Eran Malach and Samy Jelassi	644
<i>Aurora-M: Open Source Continual Pre-training for Multilingual Language and Code</i>	
Taishi Nakamura, Mayank Mishra, Simone Tedeschi, Yekun Chai, Jason T. Stillerman, Felix Friedrich, Prateek Yadav, Tanmay Laud, Vu Minh Chien, Terry Yue Zhuo, Diganta Misra, Ben Bogin, Xuan-Son Vu, Marzena Karpinska, Arnav Varma Dantuluri, Wojciech Kusa, Tommaso Furlanello, Rio Yokota, Niklas Muennighoff, Suhas Pai, Tosin Adewumi, Veronika Laippala, Xiaozhe Yao, Adalberto Barbosa Junior, Aleksandr Drozd, Jordan Clive, Kshitij Gupta, Liangyu Chen, Qi Sun, Ken Tsui, Nour Moustafa-Fahmy, Nicolo Monti, Tai Dang, Ziyang Luo, Tien-Tung Bui, Roberto Navigli, Virendra Mehta, Matthew Blumberg, Victor May, Hiep Nguyen and Sampo Pyysalo	656
<i>UCTG: A Unified Controllable Text Generation Framework for Query Auto-Completion</i>	
Zhipeng Li, Shuang Zheng, Jiaping Xiao, Xianneng Li and Lei Wang	679
<i>Lightweight Safety Guardrails Using Fine-tuned BERT Embeddings</i>	
Aaron Zheng, Mansi Rana and Andreas Stolcke	689
<i>Zero-shot Slot Filling in the Age of LLMs for Dialogue Systems</i>	
Mansi Rana, Kadri Hacioglu, Sindhuja Gopalan and Maragathamani Boothalingam	697
<i>LionGuard: A Contextualized Moderation Classifier to Tackle Localized Unsafe Content</i>	
Jessica Foo and Shaun Khoo	707

<i>REVerSum: A Multi-staged Retrieval-Augmented Generation Method to Enhance Wikipedia Tail Biographies through Personal Narratives</i>	
Sayantana Adak, Pauras Mangesh Meher, Paramita Das and Animesh Mukherjee	732
<i>RE-FIN: Retrieval-based Enrichment for Financial data</i>	
Lorenzo Malandri, Fabio Mercorio, Mario Mezzananza and Filippo Pallucchini	751
<i>SCV: Light and Effective Multi-Vector Retrieval with Sequence Compressive Vectors</i>	
Cheoneum Park, Seohyeong Jeong, Minsang Kim, KyungTae Lim and Yong-Hun Lee	760
<i>Efficient Vocabulary Reduction for Small Language Models</i>	
Yuta Nozaki, Dai Nakashima, Ryo Sato and Naoki Asaba	771
<i>Towards Boosting LLMs-driven Relevance Modeling with Progressive Retrieved Behavior-augmented Prompting</i>	
Zeyuan Chen, Haiyan Wu, Kaixin Wu, Wei Chen, Mingjie Zhong, Jia Xu, Zhongyi Liu and Wei Zhang	784
<i>LLM ContextBridge: A Hybrid Approach for Intent and Dialogue Understanding in IVSR</i>	
Changwoo Chun, Daniel Rim and Juhee Park	794
<i>Neural Document Segmentation Using Weighted Sliding Windows with Transformer Encoders</i>	
Saeed Abbasi, Aijun An, Heidar Davoudi, Ron Di Carantonio and Gary Farmaner	807
<i>RecStream: Graph-aware Stream Management for Concurrent Recommendation Model Online Serving</i>	
Shuxi Guo, Qi Qi, haifeng sun, Jianxin Liao and Jingyu Wang	817
<i>Teaching-Inspired Integrated Prompting Framework: A Novel Approach for Enhancing Reasoning in Large Language Models</i>	
Wenting Tan, Dongxiao Chen, Jieting Xue, Zihao Wang and Taijie Chen	827

Conference Program

STAND-Guard: A Small Task-Adaptive Content Moderation Model

Minjia Wang, Pingping Lin, Siqi Cai, Shengnan An, Shengjie Ma, Zeqi Lin, Congrui Huang and Bixiong Xu

Query-LIFE: Query-aware Language Image Fusion Embedding for E-Commerce Relevance

Hai Zhu, Yuankai Guo, Ronggang Dou and Kai Liu

Improving Tool Retrieval by Leveraging Large Language Models for Query Generation

Mohammad Kachuee, Sarthak Ahuja, Vaibhav Kumar, Puyang Xu and Xiaohu Liu

Know Your RAG: Dataset Taxonomy and Generation Strategies for Evaluating RAG Systems

Rafael Teixeira de Lima, Shubham Gupta, Cesar Berrospi Ramis, Lokesh Mishra, Michele Dolfi, Peter Staar and Panagiotis Vagenas

RED-CT: A Systems Design Methodology for Using LLM-labeled Data to Train and Deploy Edge Linguistic Classifiers

David Farr, Nico Manzonelli, Iain Cruickshank and Jevin West

Beyond Visual Understanding Introducing PARROT-360V for Vision Language Model Benchmarking

Harsha Vardhan Khurdula, Basem Rizk and Indus Khaitan

PDC & DM-SFT: A Road for LLM SQL Bug-Fix Enhancing

Yiwen Duan, Yonghong Yu, Xiaoming Zhao, Yichang Wu and Wenbo Liu

Multilingual Continual Learning using Attention Distillation

Sanjay Agrawal, Deep Nayak and Vivek Varadarajan Sembium

FS-DAG: Few Shot Domain Adapting Graph Networks for Visually Rich Document Understanding

Amit Agarwal, Srikant Panda and Kulbhushan Pachauri

OKG: On-the-Fly Keyword Generation in Sponsored Search Advertising

Zhao Wang, Briti Gangopadhyay, Mengjie Zhao and Shingo Takamatsu

Best Practices for Distilling Large Language Models into BERT for Web Search Ranking

Dezhi Ye, Junwei Hu, Jiabin Fan, Bowen Tian, Jie Liu, Haijin Liang and Jin Ma

Rationale-Guided Distillation for E-Commerce Relevance Classification: Bridging Large Language Models and Lightweight Cross-Encoders

Sanjay Agrawal, Faizan Ahemad and Vivek Varadarajan Sembium

Automated Clinical Data Extraction with Knowledge Conditioned LLMs

Diya Li, Asim Kadav, Aijing Gao, Rui Li and Richard Bourgon

Can Large Language Models Serve as Effective Classifiers for Hierarchical Multi-Label Classification of Scientific Documents at Industrial Scale?

Seyed Amin Tabatabaei, Sarah Fancher, Michael Parsons and Arian Askari

EDAR: A pipeline for Emotion and Dialogue Act Recognition

Elie Dina, Rania Ayachi Kibech and Miguel Couceiro

No Size Fits All: The Perils and Pitfalls of Leveraging LLMs Vary with Company Size

Ashok Urlana, Charaka Vinayak Kumar, Bala Mallikarjunarao Garlapati, Ajeet Kumar Singh and Rahul Mishra

Predicting Fine-tuned Performance on Larger Datasets Before Creating Them

Toshiki Kuramoto and Jun Suzuki

A Recipe For Building a Compliant Real Estate Chatbot

Navid Madani, Anusha Bagalkotkar, Supriya Anand, Gabriel Arnson, Rohini K. Srihari and Kenneth Joseph

Geo-Spatially Informed Models for Geocoding Unstructured Addresses

Uddeshya Singh, Devanapalli Ravi Shankar, Gowtham Bellala and Vikas Goel

Resource-Efficient Anonymization of Textual Data via Knowledge Distillation from Large Language Models

Tobias Deußer, Max Hahnbück, Tobias Uelwer, Cong Zhao, Christian Bauckhage and Rafet Sifa

Fine-Tuning Medium-Scale LLMs for Joint Intent Classification and Slot Filling: A Data-Efficient and Cost-Effective Solution for SMEs

Maia Aguirre, Ariane Méndez, Arantza del Pozo, Maria Ines Torres and Manuel Torralbo

Enhancing Large Language Models for Scientific Multimodal Summarization with Multimodal Output

Zusheng TAN, Xinyi Zhong, Jing-Yu Ji, Wei JIANG and Billy Chiu

"Stupid robot, I want to speak to a human!" User Frustration Detection in Task-Oriented Dialog Systems

Mireia Hernandez Caralt, Ivan Sekulic, Filip Carevic, Nghia Khau, Diana Nicoleta Popa, Bruna Guedes, Victor Guimaraes, Zeyu Yang, Andre Manso, Meghana Reddy, Paolo Rosso and Roland Mathis

LLM Evaluate: An Industry-Focused Evaluation Tool for Large Language Models

Harsh Saini, Md Tahmid Rahman Laskar, Cheng Chen, Elham Mohammadi and David Rossouw

Enhancing Future Link Prediction in Quantum Computing Semantic Networks through LLM-Initiated Node Features

Gilchan Park, Paul Baity, Byung-Jun Yoon and Adolfo Hoisie

Page Stream Segmentation with LLMs: Challenges and Applications in Insurance Document Automation

Hunter Heidenreich, Ratish Dalvi, Nikhil Verma and Yosheb Getachew

Graph-Augmented Open-Domain Multi-Document Summarization

Xiaoping SHEN and Yekun Chai

Improve Speech Translation Through Text Rewrite

Jing Wu, Shushu Wang, Kai Fan, Wei Luo, Minpeng Liao and Zhongqiang Huang

CarMem: Enhancing Long-Term Memory in LLM Voice Assistants through Category-Bounding

Johannes Kirmayr, Lukas Stappen, Phillip Schneider, Florian Matthes and Elisabeth Andre

XTR meets ColBERTv2: Adding ColBERTv2 Optimizations to XTR

Riyaz Ahmad Bhat and Jaydeep Sen

sDPO: Don't Use Your Data All at Once

Dahyun Kim, Yungi Kim, Wonho Song, Hyeonwoo Kim, Yunsu Kim, SANGHOON KIM and Chanjun Park

Contextual ASR Error Handling with LLMs Augmentation for Goal-Oriented Conversational AI

Yuya Asano, Sabit Hassan, Paras Sharma, Anthony B. Sicilia, Katherine Atwell, Diane Litman and Malihe Alikhani

Federated Retrieval Augmented Generation for Multi-Product Question Answering

Parshin Shojaee, Sai Sree Harsha, Dan Luo, Akash Maharaj, Tong Yu and Yunyao Li

Luna: A Lightweight Evaluation Model to Catch Language Model Hallucinations with High Accuracy and Low Cost

Masha Belyi, Robert Friel, Shuai Shao and Atindriyo Sanyal

Seeing Beyond: Enhancing Visual Question Answering with Multi-Modal Retrieval

Boqi Chen, Anuj Khare, Gaurav Kumar, Arjun Akula and Pradyumna Narayana

AutoProteinEngine: A Large Language Model Driven Agent Framework for Multi-modal AutoML in Protein Engineering

Yungeng Liu, Zan Chen, Yuguang Wang and Yiqing Shen

Building a Family of Data Augmentation Models for Low-cost LLM Fine-tuning on the Cloud

Chengyu Wang, Yuanhao Yue, Jun Huang and Peng Wang

Where do LLMs Encode the Knowledge to Assess the Ambiguity?

Hancheol Park and Geonmin Kim

On the effective transfer of knowledge from English to Hindi Wikipedia

Paramita Das, Amartya Roy, Ritabrata Chakraborty and Animesh Mukherjee

BackMATH: Towards Backward Reasoning for Solving Math Problems Step by Step

Shaowei Zhang and Deyi Xiong

Deploying Multi-task Online Server with Large Language Model

Yincen Qu, Hengyue Liu, Kun Wang, Xiangying Dai, Xiaou Lu, Hui Zhou and Chao Ma

LLM-Friendly Knowledge Representation for Customer Support

Hanchen Su, Wei Luo, Yashar Mehdad, Wei Han, Elaine Liu, Wayne Zhang, Mia Zhao and Joy Zhang

Leveraging Multilingual Models for Robust Grammatical Error Correction Across Low-Resource Languages

Divesh Ramesh Kubal and Apurva Shrikant Nagvenkar

A Simple yet Efficient Prompt Compression Method for Text Classification Data Annotation Using LLM

Yiran Xie, Debin Xiao, Ping Wang and Shuming Liu

AMAN: Agent for Mentoring and Assisting Newbies in MMORPG

Jeehyun Lee, Seung-Moo Yang and Won Ik Cho

KARRIEREWEGE: A large scale Career Path Prediction Dataset

Elena Senger, Yuri Campbell, Rob van der Goot and Barbara Plank

Transforming Code Understanding: Clustering-Based Retrieval for Improved Summarization in Domain-Specific Languages

Baban Gain, Dibyanayan Bandyopadhyay, Samrat Mukherjee, Aryan Sahoo, Saswati Dana, Palanivel Kodeswaran, Sayandeep Sen, Asif Ekbal and Dinesh Garg

Is my Meeting Summary Good? Estimating Quality with a Multi-LLM Evaluator

Frederic Thomas Kirstein, Terry Lima Ruas and Bela Gipp

Learning to Rewrite Negation Queries in Product Search

Mengtian Guo, Mutasem Al-Darabsah, Choon Hui Teo, Jonathan May, Tarun Agarwal and Rahul Bhagat

LAW: Legal Agentic Workflows for Custody and Fund Services Contracts

William Watson, Nicole Cho, Nishan Srishankar, Zhen Zeng, Lucas Cecchi, Daniel Scott, Suchetha Siddagangappa, Rachneet Kaur, Tucker Balch and Manuela Veloso

UR2N: Unified Retriever and Reranker

Riyaz Ahmad Bhat, Jaydeep Sen, Rudra Murthy and Vignesh P

An Automatic Method to Estimate Correctness of RAG

Chi Zhang, Vivek V. Datla, Aditya Shrivastava, Alf Samuel, Zhiqi Huang, Anoop Kumar and Daben Liu

DaCoM: Strategies to Construct Domain-specific Low-resource Language Machine Translation Dataset

Junghoon Kang, Keunjoo Tak, Jounghsu Choi, Myunghyun Kim, Junyoung Jang and Youjin Kang

ChartGemma: Visual Instruction-tuning for Chart Reasoning in the Wild

Ahmed Masry, Megh Thakkar, Aayush Bajaj, Aaryaman Kartha, Enamul Hoque and Shafiq Joty

LoRA Soups: Merging LoRAs for Practical Skill Composition Tasks

Akshara Prabhakar, Yuanzhi Li, Karthik Narasimhan, Sham Kakade, Eran Malach and Samy Jelassi

Aurora-M: Open Source Continual Pre-training for Multilingual Language and Code

Taishi Nakamura, Mayank Mishra, Simone Tedeschi, Yekun Chai, Jason T. Stillerman, Felix Friedrich, Prateek Yadav, Tanmay Laud, Vu Minh Chien, Terry Yue Zhuo, Diganta Misra, Ben Bogin, Xuan-Son Vu, Marzena Karpinska, Arnav Varma Dantuluri, Wojciech Kusa, Tommaso Furlanello, Rio Yokota, Niklas Muennighoff, Suhas Pai, Tosin Adewumi, Veronika Laippala, Xiaozhe Yao, Adalberto Barbosa Junior, Aleksandr Drozd, Jordan Clive, Kshitij Gupta, Liangyu Chen, Qi Sun, Ken Tsui, Nour Moustafa-Fahmy, Nicolo Monti, Tai Dang, Ziyang Luo, Tien-Tung Bui, Roberto Navigli, Virendra Mehta, Matthew Blumberg, Victor May, Hiep Nguyen and Sampo Pyysalo

UCTG: A Unified Controllable Text Generation Framework for Query Auto-Completion

Zhipeng Li, Shuang Zheng, Jiaping Xiao, Xianneng Li and Lei Wang

Lightweight Safety Guardrails Using Fine-tuned BERT Embeddings

Aaron Zheng, Mansi Rana and Andreas Stolcke

Zero-shot Slot Filling in the Age of LLMs for Dialogue Systems

Mansi Rana, Kadri Hacioglu, Sindhuja Gopalan and Maragathamani Boothalingam

LionGuard: A Contextualized Moderation Classifier to Tackle Localized Unsafe Content

Jessica Foo and Shaun Khoo

REVerSum: A Multi-staged Retrieval-Augmented Generation Method to Enhance Wikipedia Tail Biographies through Personal Narratives

Sayantan Adak, Pauras Mangesh Meher, Paramita Das and Animesh Mukherjee

RE-FIN: Retrieval-based Enrichment for Financial data

Lorenzo Malandri, Fabio Mercorio, Mario Mezzananza and Filippo Pallucchini

SCV: Light and Effective Multi-Vector Retrieval with Sequence Compressive Vectors

Cheoneum Park, Seohyeong Jeong, Minsang Kim, KyungTae Lim and Yong-Hun Lee

Efficient Vocabulary Reduction for Small Language Models

Yuta Nozaki, Dai Nakashima, Ryo Sato and Naoki Asaba

Towards Boosting LLMs-driven Relevance Modeling with Progressive Retrieved Behavior-augmented Prompting

Zeyuan Chen, Haiyan Wu, Kaixin Wu, Wei Chen, Mingjie Zhong, Jia Xu, Zhongyi Liu and Wei Zhang

LLM ContextBridge: A Hybrid Approach for Intent and Dialogue Understanding in IVSR

Changwoo Chun, Daniel Rim and Juhee Park

Neural Document Segmentation Using Weighted Sliding Windows with Transformer Encoders

Saeed Abbasi, Aijun An, Heidar Davoudi, Ron Di Carlantonio and Gary Farmaner

RecStream: Graph-aware Stream Management for Concurrent Recommendation Model Online Serving

Shuxi Guo, Qi Qi, haifeng sun, Jianxin Liao and Jingyu Wang

Teaching-Inspired Integrated Prompting Framework: A Novel Approach for Enhancing Reasoning in Large Language Models

Wenting Tan, Dongxiao Chen, Jieting Xue, Zihao Wang and Taijie Chen