



07.10.2025

Oral Talks 1 - 10:00 AM - 11:00 AM at 517BC

Fast Controlled Generation from Language Models with Adaptive Weighted Rejection Sampling

Ben Lipkin, Benjamin LeBrun, Jacob Hoover Vigly, João Loula, David R. MacIver, Lei Du, Jason Eisner, Ryan Cotterell, Vikash Mansinghka, Timothy J. O'Donnell, Alexander K. Lew, Tim Vieira

Poster Session 1 - 11:00 AM -1:00 PM at 710

BiXSE: Improving Dense Retrieval via Probabilistic Graded Relevance Distillation

Christos Tsirigotis, Vaibhav Adlakha, Joao Monteiro, Aaron C. Courville, Perouz Taslakian

Poster Session 2 - 4:30 PM - 6:30 PM at 710

Exploring Sparse Adapters for Scalable Merging of Parameter Efficient Experts

Samin Yeasar Arnob, Zhan Su, Minseon Kim, Oleksiy Ostapenko, Riyasat Ohib, Esra'a Saleh, Doina Precup, Lucas Caccia, Alessandro Sordoni

Fleurs-SLU: A Massively Multilingual Benchmark for Spoken Language Understanding

Fabian David Schmidt, Ivan Vulić, Goran Glavaš, David Ifeoluwa Adelani

Language Agents Mirror Human Causal Reasoning Biases. How Can We Help Them Think Like Scientists?

Anthony GX-Chen, Dongyan Lin, Mandana Samiei, Doina Precup, Blake Aaron Richards, Rob Fergus, Kenneth Marino

08.10.2025

Poster Session 3 - 11:00 AM-1 PM at 710

Training Plug-and-Play Knowledge Modules with Deep Context Distillation

Lucas Caccia, Alan Ansell, Edoardo Ponti, Ivan Vulić, Alessandro Sordoni

DoomArena: A framework for Testing AI Agents Against Evolving Security Threats

Léo Boisvert, Abhay Puri, Gabriel Huang, Mihir Bansal, Chandra Kiran Reddy Evuru, Avinandan Bose, Maryam Fazel, Quentin Cappart, Alexandre Lacoste, Jason Stanley, Alexandre Drouin, Krishnamurthy Dj Dvijotham

AgentRewardBench: Evaluating Automatic Evaluations of Web Agent Trajectories

Xing Han Lü, Amirhossein Kazemnejad, Nicholas Meade, Arkil Patel, Dongchan Shin, Alejandra Zambrano, Karolina Stańczak, Peter Shaw, Christopher Pal, Siva Reddy

Poster Session 4 - 4:30 PM - 6:30 PM

Putting the Value Back in RL: Better Test-Time Scaling by Unifying LLM Reasoners With Verifiers

Kusha Sareen, Morgane M Moss, Alessandro Sordoni, Arian Hosseini, Rishabh Agarwal

Steering Large Language Model Activations in Sparse Spaces

Reza Bayat, Ali Rahimi-Kalahroudi, Mohammad Pezeshki, Sarath Chandar, Pascal Vincent

Rethinking Safety in LLM Fine-tuning: An Optimization Perspective

Minseon Kim, Jin Myung Kwak, Lama Alssum, Bernard Ghanem, Philip Torr, David Krueger, Fazl Barez, Adel Bibi

BigCharts-R1: Enhanced Chart Reasoning with Visual Reinforcement Finetuning

Ahmed Masry, Abhay Puri, Masoud Hashemi, Juan A. Rodriguez, Megh Thakkar, Khyati Mahajan, Vikas Yadav, Sathwik Tejaswi Madhusudhan, Alexandre Piché, Dzmitry Bahdanau, Christopher Pal, David Vazquez, Enamul Hoque, Perouz Taslakian, Sai Rajeswar, Spandana Gella

Adaptive Computation Pruning for the Forgetting Transformer

Zhixuan Lin, Johan Samir Obando Ceron, Xu Owen He, Aaron C. Courville

Not All Data Are Unlearned Equally

Aravind Krishnan, Siva Reddy, Marius Mosbach

Do Biased Models Have Biased Thoughts?

Swati Rajwal, Shivank Garg, Reem Abdel-Salam, Abdelrahman Zayed

Resona: Improving Context Copying in Linear Recurrence Models with Retrieval

Xinyu Wang, Linrui Ma, Jerry Huang, Peng Lu, Prasanna Parthasarathi, Xiao-Wen Chang, Boxing Chen, Yufei Cui

09.10.2025

Poster Session 5 - 11:00 AM - 1:00 PM at 710

Partial Perspectives: How LLMs Handle Logically Inconsistent Knowledge in Reasoning Tasks

Zichao Li, Ines Arous, Jackie CK Cheung

Boosting LLM Reasoning via Spontaneous Self-Correction

Xutong Zhao, Tengyu Xu, Xuewei Wang, Zhengxing Chen, Di Jin, Liang Tan, Yen-Ting Lin, Zishun Yu, Zhuokai Zhao, Si-Yuan Wang, Yun He, Sinong Wang, Han Fang, Sarath Chandar, Chen Zhu