Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)

Systems Demonstrations

Miami, Florida, USA
12-16 November 2024

ISBN: 979-8-3313-0838-4

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2024) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

FreeEval: A Modular Framework for Trustworthy and Efficient Evaluation of Large Language Models
Zhuohao Yu, Chang Gao, Wenjin Yao, Yidong Wang, Zhengran Zeng, Wei Ye, Jindong Wang, Yue Zhang and Shikun Zhang 1

i-Code Studio: A Configurable and Composable Framework for Integrative AI
Yuwei Fang, Mahmoud Khademi, Chenguang Zhu, Ziyi Yang, Reid Pryzant, Yichong Xu, Yao Qian, Takuya Yoshioka, Lu Yuan, Michael Zeng and Xuedong Huang 14

Evalverse: Unified and Accessible Library for Large Language Model Evaluation
Jihoo Kim, Wonho Song, Dahyun Kim, Yunsu Kim, Yungi Kim and Chanjun Park 25

Medico: Towards Hallucination Detection and Correction with Multi-source Evidence Fusion
Xinping Zhao, Jindi Yu, Zhenyu Liu, Jifang Wang, Dongfang Li, Yibin Chen, Baotian Hu and Min Zhang 34

OpenOmni: A Collaborative Open Source Tool for Building Future-Ready Multimodal Conversational Agents
Qiang Sun, Yuanyi Luo, Sirui Li, Wenxiao Zhang and Wei Liu 46

Lighthouse: A User-Friendly Library for Reproducible Video Moment Retrieval and Highlight Detection
Taichi Nishimura, Shota Nakada, Hokuto Munakata and Tatsuya Komatsu 53

MarkLLM: An Open-Source Toolkit for LLM Watermarking
Leyi Pan, Aiwei Liu, Zhiwei He, Zitian Gao, Xuandong Zhao, Yijian LU, Binglin Zhou, Shuliang Liu, Xuming Hu, Lijie Wen, Irwin King and Philip S. Yu 61

AUTOGEN STUDIO: A No-Code Developer Tool for Building and Debugging Multi-Agent Systems
Victor Dibia, Jingya Chen, Gagan Bansal, Suff Syed, Adam Fourney, Erkang Zhu, Chi Wang and Saleema Amershi 72

TinyAgent: Function Calling at the Edge
Lutfi Eren Erdogan, Nicholas Lee, Siddharth Jha, Sehoon Kim, Ryan Tabrizi, Suhong Moon, Coleman Richard Charles Hooper, Gopala Anumanchipalli, Kurt Keutzer and Amir Gholami 80

TruthReader: Towards Trustworthy Document Assistant Chatbot with Reliable Attribution
Dongfang Li, Xinshuo Hu, Zetian Sun, Baotian Hu, Shaolin Ye, Zifei Shan, Qian Chen and Min Zhang 89

Commentator: A Code-mixed Multilingual Text Annotation Framework
Rajvee Sheth, Shubh Nisar, Heenaben Prajapati, Himanshu Beniwal and Mayank Singh 101

Integrating INCEpTION into larger annotation processes
Richard Eckart De Castilho, Jan-Christoph Klie and Iryna Gurevych 110

Arxiv Copilot: A Self-Evolving and Efficient LLM System for Personalized Academic Assistance
Guanyu Lin, Tao Feng, Pengrui Han, Ge Liu and Jiaxuan You 122

TransAgents: Build Your Translation Company with Language Agents
Minghao Wu, Jiahao Xu and Longyue Wang 131

Monitoring Hate Speech in Indonesia: An NLP-based Classification of Social Media Texts
Musa Izzanardi Wijanarko, Lucky Susanto, Prasetia Anugrah Pratama, Ika Karlina Idris, Traci Hong and Derry Tanti Wijaya 142

CAVA: A Tool for Cultural Alignment Visualization & Analysis
Nevan Giuliani, Cheng Charles Ma, Prakruthi Pradeep and Daphne Ippolito 153

ReDel: A Toolkit for LLM-Powered Recursive Multi-Agent Systems
Andrew Zhu, Liam Dugan and Chris Callison-Burch 162

BattleAgent: Multi-modal Dynamic Emulation on Historical Battles to Complement Historical Analysis
Shuhang Lin, Wenyue Hua, Lingyao Li, Che-Jui Chang, Lizhou Fan, Jianchao ji, Hang Hua, Mingyu Jin, Jiebo Luo and Yongfeng Zhang 172

sign.mt: Real-Time Multilingual Sign Language Translation Application
Amit Moryossef 182

WebOlympus: An Open Platform for Web Agents on Live Websites
Boyuan Zheng, Boyu Gou, Scott Salisbury, Zheng Du, Huan Sun and Yu Su 187

TAIL: A Toolkit for Automatic and Realistic Long-Context Large Language Model Evaluation
Gefei Gu, Yilun Zhao, Ruoxi Ning, Yanan Zheng and Arman Cohan 198

OpenResearcher: Unleashing AI for Accelerated Scientific Research
Yuxiang Zheng, Shichao Sun, Lin Qiu, Dongyu Ru, Cheng Jiayang, Xuefeng Li, Jifan Lin, Binjie Wang, Yun Luo, Renjie Pan, Yang Xu, Qingkai Min, Zizhao Zhang, Yiwen Wang, Wenjie Li and Pengfei Liu 209

OpenFactCheck: A Unified Framework for Factuality Evaluation of LLMs
Hasan Iqbal, Yuxia Wang, Minghan Wang, Georgi Nenkov Georgiev, Jiahui Geng, Iryna Gurevych and Preslav Nakov 219

ULLME: A Unified Framework for Large Language Model Embeddings with Generation-Augmented Learning
Hieu Man, Nghia Trung Ngo, Franck Dernoncourt and Thien Huu Nguyen 230

To the Globe (TTG): Towards Language-Driven Guaranteed Travel Planning
Da JU, Song Jiang, Andrew Cohen, Aaron Foss, Sasha Mitts, Arman Zharmagambetov, Brandon Amos, Xian Li, Justine T Kao, Maryam Fazel-Zarandi and Yuandong Tian 240

MATSA: Multi-Agent Table Structure Attribution
Puneet Mathur, Alexa Siu, Nedim Lipka and Tong Sun 250

OpenT2T: An Open-Source Toolkit for Table-to-Text Generation
Haowei Zhang, Shengyun Si, Yilun Zhao, Lujing Xie, Zhijian Xu, Lyuhao Chen, Linyong Nan, Pengcheng Wang, Xiangru Tang and Arman Cohan 259

ChatHF: Collecting Rich Human Feedback from Real-time Conversations
Andrew Li, Zhenduo Wang, Ethan Mendes, Duong Minh Le, Wei Xu and Alan Ritter 270

KMatrix: A Flexible Heterogeneous Knowledge Enhancement Toolkit for Large Language Model
Shun Wu, Di Wu, Kun Luo, XueYou Zhang, Jun Zhao and Kang Liu 280

Xinference: Making Large Model Serving Easy
Weizheng Lu, Lingfeng Xiong, Feng Zhang, Xuye Qin and Yueguo Chen 291

RETAIN: Interactive Tool for Regression Testing Guided LLM Migration
Tanay Dixit, Daniel Lee, Sally Fang, Sai Sree Harsha, Anirudh Sureshan, Akash V Maharaj and Yunyao Li 301

ClaimLens: Automated, Explainable Fact-Checking on Voting Claims Using Frame-Semantics
Jacob Devasier, Rishabh Mediratta, Phuong Anh Le, David Huang and Chengkai Li 311

RAGViz: Diagnose and Visualize Retrieval-Augmented Generation
Tevin Wang, Jingyuan He and Chenyan Xiong 320

PyMarian: Fast Neural Machine Translation and Evaluation in Python
Thamme Gowda, Roman Grundkiewicz, Elijah Rippeth, Matt Post and Marcin Junczys-Dowmunt 328

LLM-DetectAIve: a Tool for Fine-Grained Machine-Generated Text Detection
Mervat Abassy, Kareem Elozeiri, Alexander Aziz, Minh Ngoc Ta, Raj Vardhan Tomar, Bimarsha Adhikari, Saad El Dine Ahmed, Yuxia Wang, Osama Mohammed Afzal, Zhuohan Xie, Jonibek Mansurov, Ekaterina Artemova, Vladislav Mikhailov, Rui Xing, Jiahui Geng, Hasan Iqbal, Zain Muhammad Mujahid, Tarek Mahmoud, Akim Tsvigun, Alham Fikri Aji, Artem Shelmanov, Nizar Habash, Iryna Gurevych and Preslav Nakov 336

Translation Canvas: An Explainable Interface to Pinpoint and Analyze Translation Systems
Chinmay Dandekar, Wenda Xu, Xi Xu, Siqi Ouyang and Lei Li 344

mbrs: A Library for Minimum Bayes Risk Decoding
Hiroyuki Deguchi, Yusuke Sakai, Hidetaka Kamigaito and Taro Watanabe 351

Debug Smarter, Not Harder: AI Agents for Error Resolution in Computational Notebooks
Konstantin Grotov, Artem Borzilov, Maksim Krivobok, Timofey Bryksin and Yaroslav Zharov 363

Schema-Guided Culture-Aware Complex Event Simulation with Multi-Agent Role-Play
Sha Li, Revanth Gangi Reddy, Khanh Duy Nguyen, Qingyun Wang, Yi Fung, Chi Han, Jiawei Han, Kartik Natarajan, Clare R. Voss and Heng Ji 372

SparkRA: A Retrieval-Augmented Knowledge Service System Based on Spark Large Language Model
Dayong Wu, Jiaqi Li, Baoxin Wang, Honghong Zhao, Siyuan Xue, Yanjie Yang, Zhijun Chang, Rui Zhang, Li Qian, Bo Wang, Shijin Wang, Zhixiong Zhang and Guoping Hu 382

Generative Dictionary: Improving Language Learner Understanding with Contextual Definitions
Kai-Wen Tuan, Hai-Lun Tu and Jason S. Chang 390

WalledEval: A Comprehensive Safety Evaluation Toolkit for Large Language Models
Prannaya Gupta, Le Qi Yau, Hao Han Low, I-Shiang Lee, Hugo Maximus Lim, Yu Xin Teoh, Koh Jia Hng, Dar Win Liew, Rishabh Bhardwaj, Rajat Bhardwaj and Soujanya Poria 397

RAGLAB: A Modular and Research-Oriented Unified Framework for Retrieval-Augmented Generation
Xuanwang Zhang, Yun-Ze Song, Yidong Wang, Shuyun Tang, Xinfeng Li, Zhengran Zeng, Zhen Wu, Wei Ye, Wenyuan Xu, Yue Zhang, Xinyu Dai, Shikun Zhang and Qingsong Wen 408

AutoTrain: No-code training for state-of-the-art models
Abhishek Thakur 419

Sailor: Open Language Models for South-East Asia
Longxu Dou, Qian Liu, Guangtao Zeng, Jia Guo, Jiahui Zhou, Xin Mao, Ziqi Jin, Wei Lu and Min Lin 424

RepoAgent: An LLM-Powered Open-Source Framework for Repository-level Code Documentation Generation
Qinyu Luo, Yining Ye, Shihao Liang, Zhong Zhang, Yujia Qin, Yaxi Lu, Yesai Wu, Xin Cong, Yankai Lin, Yingli Zhang, Xiaoyin Che, Zhiyuan Liu and Maosong Sun 436

DeepPavlov 1.0: Your Gateway to Advanced NLP Models Backed by Transformers and Transfer Learning
Maksim Savkin, Anastasia Voznyuk, Fedor Ignatov, Anna Korzanova, Dmitry Karpov, Alexander Popov and Vasily Konovalov 465

Kandinsky 3: Text-to-Image Synthesis for Multifunctional Generative Framework
Arkhipkin Vladimir, Viacheslav Vasilev, Andrei Filatov, Igor Pavlov, Julia Agafonova, Nikolai Gerasimenko, Anna Averchenkova, Evelina Mironova, Bukashkin Anton, Konstantin Kulikov, Andrey Kuznetsov and Denis Dimitrov 475

MIMIR: A Customizable Agent Tuning Platform for Enhanced Scientific Applications
Xiangru Tang, Chunyuan Deng, Hanminwang Hanminwang, Haoran Wang, Yilun Zhao, Wenqi Shi, Yi Fung, Wangchunshu Zhou, Jiannan Cao, Heng Ji, Arman Cohan and Mark Gerstein 486

WildVis: Open Source Visualizer for Million-Scale Chat Logs in the Wild
Yuntian Deng, Wenting Zhao, Jack Hessel, Xiang Ren, Claire Cardie and Yejin Choi 497

Instruction-Driven Game Engine: A Poker Case Study
Hongqiu Wu, Xingyuan Liu, Yan Wang and Hai Zhao 507

LM-Interview: An Easy-to-use Smart Interviewer System via Knowledge-guided Language Model Exploitation
Hanming Li, Jifan Yu, Ruimiao Li, Zhanxin Hao, Yan Xuan, Jiaxi Yuan, Bin Xu, Juanzi Li and Zhiyuan Liu 520