

NAACL 2025

**Annual Conference of the Nations of the Americas Chapter of
the Association for Computational Linguistics**

**Proceedings of the Conference
System Demonstrations**

April 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

ISBN 979-8-89176-191-9

Table of Contents

<i>Dataverse: Open-Source ETL (Extract, Transform, Load) Pipeline for Large Language Models</i> Hyunbyung Park, Sukyung Lee, Gyoungjin Gim, Yungi Kim, Dahyun Kim and Chanjun Park . . .	1
<i>ATAIGI: An AI-Powered Multimodal Learning App Leveraging Generative Models for Low-Resource Taiwanese Hokkien</i> Yun-Hsin Chu, Shuai Zhu, Shou-Yi Hung, Bo-Ting Lin, En-Shiun Annie Lee and Richard Tzong-Han Tsai	11
<i>CLEAR-Command: Coordinated Listening, Extraction, and Allocation for Emergency Response with Large Language Models</i> Achref Doula, Bela Bohlender, Max Mühlhäuser and Alejandro Sanchez Guinea	20
<i>LM-Pub-Quiz: A Comprehensive Framework for Zero-Shot Evaluation of Relational Knowledge in Language Models</i> Max Ploner, Jacek Wiland, Sebastian Pohl and Alan Akbik	29
<i>TRACE: Real-Time Multimodal Common Ground Tracking in Situated Collaborative Dialogues</i> Hannah VanderHoeven, Brady Bhalla, Ibrahim Khebour, Austin C. Youngren, Videep Venkatesha, Mariah Bradford, Jack Fitzgerald, Carlos Mabrey, Jingxuan Tu, Yifan Zhu, Kenneth Lai, Changsoo Jung, James Pustejovsky and Nikhil Krishnaswamy	40
<i>MT-LENS: An all-in-one Toolkit for Better Machine Translation Evaluation</i> Javier García Gilabert, Carlos Escolano, Audrey Mash, Xixian Liao and Maite Melero	51
<i>A Learning-based Multi-Frame Visual Feature Framework for Real-Time Driver Fatigue Detection</i> Liang Xie and Songlin Fan	61
<i>TRUSTEVAL: A Dynamic Evaluation Toolkit on Trustworthiness of Generative Foundation Models</i> Yanbo Wang, Jiayi Ye, Siyuan Wu, Chujie Gao, Yue Huang, Xiuying Chen, Yue Zhao and Xiangliang Zhang	70
<i>AutoClean: LLMs Can Prepare Their Training Corpus</i> Xingyu Shen, Shengding Hu, Xinrong Zhang, Xu Han, Xiaojun Meng, Jiansheng Wei, Zhiyuan Liu and Maosong Sun	85
<i>SeaLLMs 3: Open Foundation and Chat Multilingual Large Language Models for Southeast Asian Languages</i> Wenxuan Zhang, Hou Pong Chan, Yiran Zhao, Mahani Aljunied, Jianyu Wang, Chaoqun Liu, Yue Deng, Zhiqiang Hu, Weiwen Xu, Yew Ken Chia, Xin Li and Lidong Bing	96
<i>Prompto: An open source library for asynchronous querying of LLM endpoints</i> Ryan Sze-Yin Chan, Federico Nanni, Angus Redlarski Williams, Edwin Brown, Liam Burke-Moore, Ed Chapman, Kate Onslow, Tvesha Sippy, Jonathan Bright and Evelina Gabasova	106
<i>ESPnet-SpeechLM: An Open Speech Language Model Toolkit</i> Jinchuan Tian, Jiatong Shi, William Chen, Siddhant Arora, Yoshiki Masuyama, Takashi Maekaku, Yihan Wu, Junyi Peng, Shikhar Bharadwaj, Yiwen Zhao, Samuele Cornell, Yifan Peng, Xiang Yue, Chao-Han Huck Yang, Graham Neubig and Shinji Watanabe	116
<i>InspectorRAGet: An Introspection Platform for RAG Evaluation</i> Kshitij P Fadnis, Siva Sankalp Patel, Odellia Boni, Yannis Katsis, Sara Rosenthal, Benjamin Sznajder and Marina Danilevsky	125

<i>Cerebrum (AIOS SDK): A Platform for Agent Development, Deployment, Distribution, and Discovery</i> Balaji Rama, Kai Mei and Yongfeng Zhang	135
<i>GenSim: A General Social Simulation Platform with Large Language Model based Agents</i> Jiakai Tang, Heyang Gao, Xuchen Pan, Lei Wang, Haoran Tan, Dawei Gao, Yushuo Chen, Xu Chen, Yankai Lin, Yaliang Li, Bolin Ding, Jingren Zhou, Jun Wang and Ji-Rong Wen	143
<i>Semi-automatic Sequential Sentence Classification in the Discourse Analysis Tool Suite</i> Tim Fischer and Chris Biemann	151
<i>CowPilot: A Framework for Autonomous and Human-Agent Collaborative Web Navigation</i> Faria Huq, Zora Zhiruo Wang, Frank F. Xu, Tianyue Ou, Shuyan Zhou, Jeffrey P. Bigham and Graham Neubig	163
<i>eRevise+RF: A Writing Evaluation System for Assessing Student Essay Revisions and Providing Formative Feedback</i> Zhexiong Liu, Diane Litman, Elaine L Wang, Tianwen Li, Mason Gobat, Lindsay Clare Matsu- mura and Richard Correnti	173
<i>VERSA: A Versatile Evaluation Toolkit for Speech, Audio, and Music</i> Jiatong Shi, Hye-jin Shim, Jinchuan Tian, Siddhant Arora, Haibin Wu, Darius Petermann, Jia Qi Yip, You Zhang, Yuxun Tang, Wangyou Zhang, Dareen Safar Alharthi, Yichen Huang, Koichi Saito, Jionghao Han, Yiwen Zhao, Chris Donahue and Shinji Watanabe	191
<i>Persona-SQ: A Personalized Suggested Question Generation Framework For Real-world Documents</i> Zihao Lin, Zichao Wang, Yuanting Pan, Varun Manjunatha, Ryan A. Rossi, Angela Lau, Lifu Huang and Tong Sun	210
<i>ESPnet-SDS: Unified Toolkit and Demo for Spoken Dialogue Systems</i> Siddhant Arora, Yifan Peng, Jiatong Shi, Jinchuan Tian, William Chen, Shikhar Bharadwaj, Haya- to Futami, Yosuke Kashiwagi, Emiru Tsunoo, Shuichiro Shimizu, Vaibhav Srivastav and Shinji Wata- nabe	248
<i>SURF: A System to Unveil Explainable Risk Relations between Firms</i> Yu-Hsiang Wang, Wei-Ning Chiu, Yi-Tai Hsiao, Yu-Shiang Huang, Yi-Shyuan Chiang, Shuo-En Wu and Chuan-Ju Wang	260
<i>Libra-Leaderboard: Towards Responsible AI through a Balanced Leaderboard of Safety and Capabili- ty</i> Haonan Li, Xudong Han, Zenan Zhai, Honglin Mu, Hao Wang, Zhenxuan Zhang, Yilin Geng, Shom Lin, Renxi Wang, Artem Shelmanov, Xiangyu Qi, Yuxia Wang, Donghai Hong, Youliang Yuan, Meng Chen, Haoqin Tu, Fajri Koto, Cong Zeng, Tatsuki Kuribayashi, Rishabh Bhardwaj, Bingchen Zhao, Yawen Duan, Yi Liu, Emad A. Alghamdi, Yaodong Yang, Yinpeng Dong, Soujanya Poria, Peng- fei Liu, Zhengzhong Liu, Hector Xuguang Ren, Eduard Hovy, Iryna Gurevych, Preslav Nakov, Monojit Choudhury and Timothy Baldwin	268
<i>Unlocking Korean Verbs: A User-Friendly Exploration into the Verb Lexicon</i> Seohyun Song, Eunkyul Leah Jo, Yige Chen, Jeon-Pyo Hong, Kyuwon Kim, Jin Wee, Kang Miyong, KyungTae Lim, Jungyeul Park and Chulwoo Park	287
<i>TransformerRanker: A Tool for Efficiently Finding the Best-Suited Language Models for Downstream Classification Tasks</i> Lukas Garbas, Max Ploner and Alan Akbik	295
<i>Learning Low-Resource Languages Through NLP-Driven Flashcards: A Case Study of Hokkien in Language Learning Applications</i>	

Tai Zhang, Lucie Yang, Erin Chen, Karen Riani, Jessica Zipf, Mariana Shimabukuro and En-Shiun Annie Lee	303
<i>A Sentence-Level Visualization of Attention in Large Language Models</i>	
Seongbum Seo, Sangbong Yoo, Hyelim Lee, Yun Jang, Ji Hwan Park and Jeong-Nam Kim ..	313
<i>NeMo-Inspector: A Visualization Tool for LLM Generation Analysis</i>	
Daria Gitman, Igor Gitman and Evelina Bakhturina.....	321
<i>Cognitive Kernel: An Open-source Agent System towards Generalist Autopilots</i>	
Hongming Zhang, Xiaoman Pan, Hongwei Wang, Kaixin Ma, Wenhao Yu and Dong Yu	328
<i>SOTOPIA-S4: a user-friendly system for flexible, customizable, and large-scale social simulation</i>	
Xuhui Zhou, Zhe Su, Sophie Feng, Jiaxu Zhou, Jen-tse Huang, Hsien-Te Kao, Spencer Lynch, Svitlana Volkova, Tongshuang Wu, Anita Woolley, Hao Zhu and Maarten Sap.....	350
<i>SafeSpeech: A Comprehensive and Interactive Tool for Analysing Sexist and Abusive Language in Conversations</i>	
Xingwei Tan, Chen Lyu, Hafiz Muhammad Umer, Sahrish Khan, Mahathi Parvatham, Lois Arthurs, Simon Cullen, Shelley Wilson, Arshad Jhumka and Gabriele Pergola	361
<i>ALOHA: Empowering Multilingual Agent for University Orientation with Hierarchical Retrieval</i>	
Mingxu Tao, Bowen Tang, Mingxuan Ma, Yining Zhang, Hourun Li, Feifan Wen, Ma Hao and Jia Yang.....	383
<i>MeKB-Sim: Personal Knowledge Base-Powered Multi-Agent Simulation</i>	
Zhenran Xu, Jifang Wang, Baotian Hu, Longyue Wang and Min Zhang	393
<i>MetaScientist: A Human-AI Synergistic Framework for Automated Mechanical Metamaterial Design</i>	
Jingyuan Qi, Zian Jia, Minqian Liu, Wangzhi Zhan, Junkai Zhang, Xiaofei Wen, Jingru Gan, Jianpeng Chen, Qin Liu, Mingyu Derek Ma, Bangzheng Li, Haohui Wang, Adithya Kulkarni, Muhao Chen, Dawei Zhou, Ling Li, Wei Wang and Lifu Huang	404
<i>FACTS&EVIDENCE: An Interactive Tool for Transparent Fine-Grained Factual Verification of Machine-Generated Text</i>	
Varich Boonsanong, Vidhisha Balachandran, Xiaochuang Han, Shangbin Feng, Lucy Lu Wang and Yulia Tsvetkov	437
<i>LiteWebAgent: The Open-Source Suite for VLM-Based Web-Agent Applications</i>	
Danqing Zhang, Balaji Rama, Jingyi Ni, Shiyang He, Fu Zhao, Kunyu Chen, Arnold Chen and Junyu Cao.....	449
<i>L3GO: Language Agents with Chain-of-3D-Thoughts for Generating Unconventional Objects</i>	
Yutaro Yamada, Khyathi Chandu, Bill Yuchen Lin, Jack Hessel, Ilker Yildirim and Yejin Choi	456
<i>Constructing Multimodal Datasets from Scratch for Rapid Development of a Japanese Visual Language Model</i>	
Keito Sasagawa, Koki Maeda, Issa Sugiura, Shuhei Kurita, Naoaki Okazaki and Daisuke Kawahara.....	470
<i>Storybranch - generating multimedia content from novels</i>	
Rushikesh Hiray and Venelin Kovatchev	485
<i>EventFull: Complete and Consistent Event Relation Annotation</i>	
Alon Eirew, Eviatar Nachshoni, Aviv Slobodkin and Ido Dagan.....	494

<i>METAPHORSHARE: A Dynamic Collaborative Repository of Open Metaphor Datasets</i> Joanne Boisson, Arif Mehmood and Jose Camacho-Collados	509
<i>Towards Unified, Dynamic and Annotation-based Visualisations and Exploration of Annotated Big Data Corpora with the Help of Unified Corpus Explorer</i> Kevin Bönisch, Giuseppe Abrami and Alexander Mehler	522
<i>MobA: Multifaceted Memory-Enhanced Adaptive Planning for Efficient Mobile Task Automation</i> Zichen Zhu, Hao Tang, Yansi Li, Dingye Liu, Hongshen Xu, Kunyao Lan, Danyang Zhang, Yixuan Jiang, Hao Zhou, Chenrun Wang, Situo Zhang, Liangtai Sun, Yixiao Wang, Yuheng Sun, Lu Chen and Kai Yu	535
<i>OpenReviewer: A Specialized Large Language Model for Generating Critical Scientific Paper Reviews</i> Maximilian Idahl and Zahra Ahmadi	550