EMNLP 2023

The 2023 Conference on Empirical Methods in Natural Language Processing

Proceedings of the Industry Track

December 6-10, 2023

©2023 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 978-8-89176-068-4

Organizing Committee

Industry Track Chairs

Mingxuan Wang, ByteDance AI Lab
Imed Zitouni, Google

Program Committee

Program Committee

Ahmed Abdelali, Qatar Computing Research Institute
Asad Abdi, University of Derby
Kaori Abe, RIKEN
Sallam Abualhaija, SnT, University of Luxembourg
Jiban Adhikary, Best Buy Co., Inc
Suman Adhya, Indian Association for the Cultivation of Science
Karan Aggarwal, Amazon.com Inc
Sarthak Ahuja, Amazon Alexa AI
Mousumi Akter, Auburn University
Georgios Alexandridis, National Technical University of Athens
Israa Alghanmi, University of Jeddah
Abdulaziz Alhamadani, Virginia Tech
Kenneth Alperin, MIT
Duygu Altinok, Senior NLP Engineer
Haozhe An, University of Maryland, College Park
Raviteja Anantha, Apple
Rafael Anchiêta, Federal Institute of Piauí
Diego Antognini, IBM Research
Mario Aragon, Universidade de Santiago de Compostela
Eiji Aramaki, NAIST, Japan
Ekaterina Artemova, Toloka.AI
Ai Ti Aw, Institute for Infocomm Research
Vikas Bahirwani, Google
Long Bai, CAS Key Laboratory of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences
Mithun Balakrishna, Morgan Stanley
Debayan Banerjee, Language Technology Group, University of Hamburg
Yuwei Bao, University of Michigan
Ian Beaver, Verint Systems Inc
Dorothee Beermann, Norwegian University of Science and Technology
Gabriel Bernier-Colborne, National Research Council Canada
Dario Bertero, NLPatent
Rishabh Bhardwaj, Singapore University of Technology and Design
Mukul Bhutani, Carnegie Mellon University, Apple, Google
Nikita Bhutani, Megagon Labs
Debmalya Biswas, Wipro
Yonatan Bitton, The Hebrew University of Jerusalem
Carlos Bobed Lisbona, University of Zaragoza, Spain
Nadav Borenstein, Department of Computer Science, University of Copenhagen
Nadjet Bouayad-Agha, NLP Scientist
Daniel Braun, University of Twente
Thomas Brovelli (Meyer), Google LLC
Hannah Brown, National University of Singapore
Yi Bu, Peking University
Sky CH-Wang, Columbia University
Ruken Cakici, METU, Ankara

Fabio Casati, Servicenow and U of trento
Dumitru-Clementin Cercel, University Politehnica of Bucharest
Yekun Chai, Baidu
Yung-Chun Chang, Graduate Institute of Data Science, Taipei Medical University
Deli Chen, Tencent Inc.
Fuxiang Chen, University of Leicester
Guanhua Chen, Southern University of Science and Technology
Hanjie Chen, University of Virginia
Huadong Chen, ByteDance
Lihu Chen, Telecom Paris & Institut Polytechnique de Paris
Lin Chen, Engineering Manager, Meta Platform Inc.
Maximillian Chen, Columbia University
Tongfei Chen, Microsoft
Xingran Chen, University of Michigan
Yubo Chen, Department of Electronic Engineering, Tsinghua University
Zhiyu Chen, Amazon
Daixuan Cheng, Microsoft
Joe Cheri, NIT Calicut
Jianfeng Chi, Meta AI
Sangwoo Cho, Tecent AI Lab
Shamil Chollampatt, Zoom
Hyunseung Chung, KAIST
Bonaventura Coppola, University of Trento
Peng Cui, ETH Zurich
Bosheng Ding, Nanyang Technological University
Daniel Dahlmeier, SAP
Daniel Dakota, Indiana University
Aswarth Abhilash Dara, Walmart
Souvik Das, University at Buffalo
Steve DeNeefe, RWS Language Weaver
Jean-Benoit Delbrouck, Stanford University
Daryna Dementieva, Technical University of Munich
Shijian Deng, University of Texas at Dallas
Shumin Deng, National University of Singapore
Yifan Deng, University of Chinese Academy of Sciences
Jan Deriu, Zurich University of Applied Sciences
Prajit Dhar, University of Groningen
Dimitar Dimitrov, University of Sofia St. Kliment Ohridski"
Yuning Ding, FernUniversität in Hagen
Rahul Divekar, Educational Testing Service
Sumanth Doddapaneni, Indian Institute of Technology Madras
Bin Dong, Ricoh Software Research Center Beijing
Li Dong, Amazon.com
Ming Dong, School of Computer, Central China Normal University
Zi-Yi Dou, UCLA
Jad Doughman, Mohamed bin Zayed University of Artificial Intelligence
Eduard Dragut, Temple University
Andrew Drozdov, UMass Amherst
Sourav Dutta, Huawei Research Centre
Chris Chinenye Emezue, Technical University of Munich
Run-Ze Fan, Institute of Computing Technology(ICT) of Chinese Academy of Sciences(CAS)

Qingkai Fang, Institute of Computing Technology, Chinese Academy of Sciences
Tianqing Fang, Hong Kong University of Science and Technology
Yihao Fang, Queen's University
Zheng Fang, University of Warwick
Ruitao Feng, Saarland University
Michael Flor, Educational Testing Service
Marco Aurelio Fonseca, University of Illinois
Simona Frenda, Università degli Studi di Torino
Ramiro H. Galvez, Universidad Torcuato Di Tella
Yujian Gan, Queen Mary University of London
Krishna Garg, University of Illinois at Chicago
Xiou Ge, University of Southern California
Sahar Ghannay, CNRS, LISN
Mozhdeh Gheini, University of Southern California
Sucheta Ghosh, HITS gGmbH
Lee Gillam, University of Surrey
Voula Giouli, Institute for Language & Speech Processing, ATHENA Research & Innovation Centre
Pranav Goel, Bloomberg L.P.
Jiaying Gong, Virginia Polytechnic Institute and State University
Beliz Gunel, Google AI
Ruohao Guo, Georgia Institute of Technology
Shaoru Guo, Institute of Automation, Chinese Academy of Sciences
Tong Guo, Meituan
Joonghyuk Hahn, Computer Science, Yonsei University
Shiyi Han, Beihang University
Zhen Han, Amazon
Sabit Hassan, University of Pittsburgh
Lisa Anne Hendricks, DeepMind
Namgyu Ho, KAIST
Pengyu Hong, Brandeis University
Sho Hoshino, CyberAgent, Inc.
Wenjun Hou, The Hong Kong Polytechnic University
Yifan Hou, ETH Zurich
Phillip Howard, Intel Labs
Kristen Howell, LivePerson Inc.
Phu Mon Htut, AWS AI Labs
Guimin Hu, Harbin Institute of Technology, Shenzhen
Wei Hu, Nanjing University
Zhiyuan Hu, National University of Singapore
Jie Huang, University of Illinois at Urbana-Champaign
Zhiqi Huang, Tencent
John Hudzina, Thomson Reuters
Dae Yon Hwang, Amazon
Wonseok Hwang, LBox
Joseph Marvin Imperial, University of Bath
Koji Inoue, Kyoto University
Hayate Iso, Megagon Labs
Takumi Ito, Tohoku University / Langsmith Inc. / Utrecht University
Tomoya Iwakura, Fujitsu
Kenichi Iwatsuki, Mirai Translate Incorporated

Milos Jakubicek, Lexical Computing
Jiyue Jiang, The University of Hong Kong
Li Jin, Aerospace Information Research Institute, Chinese Academy of Sciences
Qiao Jin, National Institutes of Health
Yingnan Ju, Inidiana University
Oren Kalinsky, Amazon
Hiroshi Kanayama, IBM Research - Tokyo
Pinar Karagoz, METU Computer Eng. Dept.
Payam Karisani, UIUC
Pride Kavumba, Tohoku University / RIKEN AIP
Ashkan Kazemi, University of Michigan
Amr Keleg, The University of Edinburgh
Roman Kern, Graz University of Technology
Jiwoo Kim, Sungkyunkwan University
Takyoung Kim, LG AI Research
Tracy Holloway King, Adobe Inc.
Miyoung Ko, KAIST
Thomas Kober, Zalando SE
Jan Kocon, Wrocław University of Science and Technology
Svetla Koeva, Institute for Bulgarian Language "Prof. Lyubomir Andreychin", Bulgarian Academy of Sciences
Ana Kotarcic, University of Zurich
Vasiliki Kougia, University of Vienna
Da Kuang, Huawei
Marek Kubis, Adam Mickiewicz University
Andrei Kucharavy, HES-SO Valais-Wallis
Sanjeev Kumar, Quark.ai
Tsung-Ting Kuo, University of California San Diego
Sunjun Kweon, KAIST
Yanis Labrak, Laboratoire Informatique d'Avignon (LIA)
Kushal Lakhotia, Facebook AI Research
Wai Lam, The Chinese University of Hong Kong
Lukas Lange, Bosch Center for Artificial Intelligence
Mateusz Lango, Poznan University of Technology
Stefan Larson, Vanderbilt University
Léo Laugier, Swiss Federal Institute of Technology of Lausanne
Alexandra Lavrentovich, Amazon Alexa
Grandee Lee, National University of Singapore
Gyubok Lee, KAIST
Minwoo Lee, Seoul National University
Arun Balajiee Lekshmi Narayanan, University of Pittsburgh
Jens Lemmens, University of Antwerp
Sicong Leng, Nanyang Technological University
Yves Lepage, Waseda University
Ran Levy, Amazon
Changmao Li, UC Santa Cruz
Chong Li, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
Dingcheng Li, Coupang Search and Discovery
Dongfang Li, Harbin Institute of Technology, Shenzhen
Haochen Li, Nanyang Technological University

Jiangnan Li, Institute of Information Engineering, Chinese Academy of Sciences
Jiazhao Li, University of Michigan
Juanhui Li, Michigan State University
Keyi Li, Rutgers University
Li Erran Li, Amazon
Mengze Li, Zhejiang University
Mingchen Li, University of Minnesota Twin Cities
Mingda Li, University of California, Los Angeles
Songlin Li, Stanford University
Wei Li, Nanyang Technological University
Yinghui Li, Tsinghua University
Yingya Li, Harvard Medical School and Boston Children's Hospital
Zhigen Li, Ping An Technology
Davis Liang, Amazon
Zhengzhong Liang, Google
Veronica Liesaputra, University of Otago
Gilbert Lim, SingHealth
Nut Limsopatham, Microsoft AI+R
Lucy Lin, Spotify
Peiqin Lin, LMU Munich
Ting-En Lin, Alibaba Group
Zhenxi Lin, Tencent
Guangliang Liu, Michigan State University
Lei Liu, Department of Electrical Engineering and Computer Science, York University
Pinxin Liu, University of Rochester
Yang Janet Liu, Georgetown University
Ye Liu, Mercedes-Benz AG
Yonghao Liu, Jilin University
Yongkang Liu, Northeastern University;lmu
Zixuan Liu, University of Washington
Abhay Lokesh Kashyap, Walmart
Petr Lorenc, Czech Technical University in Prague
Natalia Loukachevitch, Lomonosov Moscow State University
Jianqiao Lu, The University of Hong Kong
Xiaolei Lu, City University of Hong Kong
Xuesong Lu, East China Normal University
Haozheng Luo, Northwestern University
Wencan Luo, Google
Ziyang Luo, Hong Kong Baptist University
Zhixin MA, Singapore Management University
Nianzu Ma, University of Minnesota Twin Cities
Xinyin Ma, National University of Singapore
Ziqiao Ma, University of Michigan
Andrew Mackey, University of Arkansas
Ayush Maheshwari, IIT Bombay
Quan Mai, University of Arkansas
Wolfgang Maier, Mercedes-Benz AG
Lorenzo Malandri, University of Milan - Bicocca
Pranav Maneriker, The Ohio State University
Kelong Mao, Gaoling School of Artificial Intelligence, Renmin University of China
Wenji Mao, Chinese Academy of Sciences

Zhiming Mao, The Chinese University of Hong Kong
Piotr Mardziel, Independent
Antonis Maronikolakis, Ludwig-Maximillians-University of Munich
Eugenio Martínez Cámara, University of Jaén
Puneet Mathur, University of Maryland College Park
Alexander Mehler, Goethe-University Frankfurt am Main
Mahnoosh Mehrabani, Interactions LLC
Telmo Menezes, Centre Marc Bloch / CNRS / Humboldt Universität
Yuanliang Meng, Nuance Communications Inc.
Eleni Metheniti, University of Toulouse 3, IRIT
Hideya Mino, NHK Science & Technology Research Laboratories
Masato Mita, CyberAgent Inc.
Hosein Mohebbi, Tilburg University
Jihyung Moon, SoftlyAI
Lori Moon, Elemental Cognition
Jose G. Moreno, Paul Sabatier University - IRIT
Makoto Morishita, NTT Communication Science Laboratories
Larry Moss, Indiana University, Bloomington
Aldrian Obaja Muis, None
Matthew Mulholland, Educational Testing Service
Vitobha Munigala, Research Engineer, IBM Research
Emir Munoz, Genesys Cloud Services, Inc.
Masayasu Muraoka, IBM Research - Tokyo
Emmanuel Ngue Um, University of Yaoundé I
Farah Nadeem, World Bank
Varun Nagaraj Rao, Princeton University
Masaaki Nagata, NTT Corporation
Tetsuji Nakagawa, Google Japan G.K.
Sungjin Nam, ACT, Inc
Marcin Namysl, Fraunhofer IAIS
Tarek Naous, Georgia Institute of Technology
Diane Napolitano, The Washington Post
Tapas Nayak, TCS Research
Hoang Nguyen, University of Illinois at Chicago
Kiet Nguyen, University of Information Technology, VNU-HCM
Jingwei Ni, ETH Zurich
Minheng Ni, Microsoft Research
Shiwen Ni, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
Shaoliang Nie, Meta Inc
Dmitry Nikolaev, University of Stuttgart
Iftitahu Nimah, Eindhoven University of Technology, BRIN Indonesia
Navid Nobani, University of Milano-Bicocca
Alexander O'Connor, Autodesk
Oleg Okun, Enpal GmbH
Eda Okur, Intel Labs
Naoki Otani, Megagon Labs
Ankur Padia, Philips Research North America
Vardaan Pahuja, The Ohio State University
Pierre-Henri Paris, Telecom Paris
Lucas Pavanelli, aiXplain
Vera Pavlova, rttl.ai

Zhengqi Pei, Chinese Academy of Sciences
Olga Pelloni, University of Zurich, Telepathy Labs
Weirui Peng, Columbia University
Fred Philippy, Zortify Labs, Zortify S.A.
Jakub Piskorski, Polish Academy of Sciences
Laura Plaza, UNED
Brian Pluss, University of Dundee
Animesh Prasad, Roku
Radityo Eko Prasojo, Pitik.id
Dongqi Pu, Saarland University
Lihua Qian, ByteDance AI Lab
Yushan Qian, Tianjin University
Xinying Qiu, School of Information Science and Technology, Guangdong University of Foreign Studies
Kanagasabai Rajaraman, Institute for Infocomm Research, A*STAR
Dongning Rao, Guangdong University of Technology
Traian Rebedea, University Politehnica of Bucharest & NVIDIA
Ryokan Ri, LINE Corporation
Matiss Rikters, AIST
Tharathorn Rimchala, Intuit Inc
Md Rashad Al Hasan Rony, University of Bonn
Kevin Ros, University of Illinois at Urbana-Champaign
Mozhdeh Rouhsedaghat, USC
Shamik Roy, Purdue University
Sumegh Roychowdhury, Indian Institute of Technology, Kharagpur
Alsu Sagirova, DeepPavlov
Monjoy Saha, NCI, NIH
Sougata Saha, State University of New York at Buffalo
Tanay Kumar Saha, Walmart Global Tech
Tanja Samardžić, University of Zurich
Minoru Sasaki, Ibaraki University
Robin Schmidt, Apple
Claudia Schulz, Thomson Reuters
Ronald Seoh, Purdue University
Sofia Serrano, University of Washington
Agam Shah, Georgia Institute of Technology
Deven Santosh Shah, Microsoft
Serge Sharoff, University of Leeds
Qinlan Shen, Oracle
Tianhao Shen, Tianjin University
Qiang Sheng, Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences
Jihao Shi, Harbin Institute of Technology
Zhouxing Shi, UCLA
Melanie Siegel, Hochschule Darmstadt - University of Applied Sciences
Ingo Siegert, Otto von Guericke University Magdeburg
Andrew Silva, Georgia Institute of Technology
Patrick Simianer, Lilt
Sneha Singhania, Max Planck Institute for Informatics
Priyanka Sinha, Independent
Victor Skobov, The Graduate University for Advanced Studies (SOKENDAI)

Ilia Sominsky, Datto
Seonil (Simon) Son, NLP Center, NCSOFT
Hyun-Je Song, Jeonbuk National University
Yan Song, USTC
Anna Sotnikova, University of Maryland
Makesh Narsimhan Sreedhar, University of Wisconsin-Madison
Mukund Srinath, Pennsylvania State University
Katherine Stasaski, Salesforce Research
Michal Stefanik, Masaryk University
Evgeny Stepanov, VUI, Inc.
Ian Stewart, Pacific Northwest National Laboratory
Kristina Striegnitz, Union College
David Strohmaier, University of Cambridge
Sebastian Stüker, Zoom Video Communications Inc.
Ying Su, HKUST
Yoshi Suhara, Grammarly
Chengjie Sun, Harbin Institute of Technology
Chenkai Sun, University of Illinois at Urbana-Champaign
Qiushi Sun, East China Normal University
Weixuan Sun, Australian National University
Marek Suppa, Comenius University in Bratislava
Munira Syed, Procter & Gamble
Santosh T.Y.S.S, Technical University of Munich
Bilal Taha, University of Toronto
Dima Taji, Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics
Yu Takagi, Osaka University
Chunxu Tang, Twitter
Raphael Tang, Comcast
Tianyi Tang, Renmin University of China
Xuemei Tang, Peking University
Yun Tang, Facebook
Yuanhe Tian, Department of Linguistics, University of Washington
Manabu Torii, Kaiser Permanente Southern California
Lucas Torroba Hennigen, Massachusetts Institute of Technology
Benjamin Towle, University of Nottingham
Masaaki Tsuchida, COTOBA DESIGN, Inc.
Adrian Ulges, RheinMain University of Applied Sciences
David Uthus, Google Research
Nidhi Vakil, University of Massachusetts, Lowell
Daniel Varab, Novo Nordisk & IT University of Copenhagen
Manuel Vilares Ferro, University of Vigo
Bingqing Wang, Bosch Research & Technology Center North America
Chengyu Wang, Alibaba Group
Feifei Wang, Renmin University of China
Hai Wang, Toyota Technological Institute at Chicago
Huimin Wang, Tencent
Jianzong Wang, Ping An Technology (Shenzhen) Co., Ltd.
Jie Wang, Southwest jiaotong university
Jun Wang, iWudao Tech
Jun Wang, Ludong University

Runze Wang, Alibaba Group
Ryan Wang, University of Illinois Urbana-Champaign
Suge Wang, School of Computer & Information Technology, Shanxi University
Weiqi Wang, Hong Kong University of Science and Technology
Yanshan Wang, University of Pittsburgh
Yile Wang, Tsinghua University
Yiran Wang, National Institute of Information and Communications Technology
Zhilin Wang, Nvidia
Zihao Wang, HKUST
Penghui Wei, Alibaba Group
Di Wu, University of California, Los Angeles
Junjie Wu, Hong Kong University of Science and Technology
Tianxing Wu, Southeast University
Ting Wu, Fudan University
Tingting Wu, Harbin Institute of Technology
Yuxia Wu, Xi'an Jiaotong University
Zhuofeng Wu, University of Michigan
Yuqing Xie, The University of Waterloo
Hangdi Xing, Zhejiang University
Chunyang Xiao, Amazon Alexa
Min Xiao, Microsoft
Huanyi Xie
Kaige Xie, Georgia Institute of Technology
Hongyan Xu
Jinan Xu, Beijing Jiaotong University
Jitao Xu, NetEase Youdao, Tsinghua University
Jun Xu, Baidu
Mingzhou Xu, University of Macau;
Wang Xu, Harbin Institute of Technology
Yi Xu, Shanghai Jiao Tong University
Dejie Yang, Peking University
Shu Yang, University of Macau
Wei Yang, Wish
Zhichao Yang, University of Massachusetts
Zi Yang, Google
Liang Yao, Tencent
Bingyang Ye, Brandeis University
Fanghua Ye, University College London
Rong Ye, ByteDance AI Lab
Wenting Ye, ByteDance Inc
Mohammad Yeghaneh Abkenar, Phd Candidate Reseacher
Jinyoung Yeo, Yonsei University
Jingwei Yi, University of Science and Technology of China
Xuwang Yin, University of Virginia
Yuwei Yin, The University of British Columbia
Susik Yoon, University of Illinois at Urbana-Champaign
Issei Yoshida, IBM Research - Tokyo
Cheng Yu, ShanghaiTech University
Lei Yu, Sino-French Engineer School of Beihang University and Beihang Hangzhou Innovation Institute Yuhang
Xinchen Yu, University of Arizona

Zac Yu, Google
Siyu Yuan, Fudan University
Zheng Yuan, Italian Institute of Technology
Mahdi Zakizadeh, Tehran Institute for Advanced Studies
Qi Zeng, University of Illinois at Urbana-Champaign
Qingkai Zeng, University of Notre Dame
Yawen Zeng, ByteDance AI Lab
Runzhe Zhan, University of Macau
Baohua Zhang, Beijing Institute of Technology
Chen Zhang, ECE, National University of SIngapore
Dan Zhang, Megagon Labs
Lei Zhang, LinkedIn
Shaolei Zhang, Institute of Computing Technology, Chinese Academy of Sciences
Shuo Zhang, Xi'an Jiaotong University, P. R. China
Tianlin Zhang, The University of Manchester
Weixu Zhang, Xi'an Jiaotong University
Yin Zhang, Google
Yu Zhang, Google Brain
Zhe Zhang, Meta AI
Zhehao Zhang, Shanghai Jiao Tong University
Chao Zhao, University of North Carolina at Chapel Hill
Jiahao Zhao, Institute of Automation, Chinese Academy of Sciences
Xuandong Zhao, UC Santa Barbara
Changmeng Zheng, The Hong Kong Polytechnic University
Junhao Zheng, South China University of Technology
Zi'ou Zheng, Queen's University
Baohang Zhou, Nankai University
Dong Zhou, Guangdong University of Foreign Studies
Junpei Zhou, Apple
Yu Zhou, University of California, Los Angeles
Zhengyu Zhou, Bosch Research and Technology Center North America
Luyao Zhu, Nanyang Technological University
Su Zhu, AISpeech Co., Ltd
Xiliang Zhu, Dialpad
Xuan Zhu, Amazon
Xuekai Zhu, Department of Computer Science and Engineering, Shanghai Jiao Tong University
Jin Ziqi, Singapore University of Technology and Design
Bowei Zou, Institute for Infocomm Research
Amal Zouaq, Polytechnique Montreal

Table of Contents

BeautifulPrompt: Towards Automatic Prompt Engineering for Text-to-Image Synthesis
Tingfeng Cao, Chengyu Wang, Bingyan Liu, Ziheng Wu, Jinhui Zhu and Jun Huang 1

Enhancing Language Model with Unit Test Techniques for Efficient Regular Expression Generation
Chenhui Mao, Xiexiong Lin, Xin Jin and Xin Zhang 12

A Comparative Analysis of Task-Agnostic Distillation Methods for Compressing Transformer Language Models
Takuma Udagawa, Aashka Trivedi, Michele Merler and Bishwaranjan Bhattacharjee 20

Towards Effective Automatic Debt Collection with Persona Awareness
Tong Zhang, Junhong Liu, Chen Huang, Jia Liu, Hongru Liang, Zujie Wen and Wenqiang Lei 32

Gatekeeper to save COGS and improve efficiency of Text Prediction
Nidhi Tiwari, Sneha Kola, Milos Milunovic, Si-qing Chen and Marjan Slavkovski 46

Efficient Transformer Knowledge Distillation: A Performance Review
Nathan Brown, Ashton Williamson, Tahj Anderson and Logan Lawrence 54

CDD: A Large Scale Dataset for Legal Intelligence Research
Changzhen Ji, Yating Zhang, Adam Jatowt and Haipang Wu 66

MUST&P-SRL: Multi-lingual and Unified Syllabification in Text and Phonetic Domains for Speech Representation Learning
Noé Tits 74

Personalized Dense Retrieval on Global Index for Voice-enabled Conversational Systems
Masha Belyi, Charlotte Dzialo, Chaitanya Dwivedi, Prajit Reddy Muppidi and Kanna Shimizu 83

Text2Topic: Multi-Label Text Classification System for Efficient Topic Detection in User Generated Content with Zero-Shot Capabilities
Fengjun Wang, Moran Beladev, Ofri Kleinfeld, Elina Frayerman, Tal Shachar, Eran Fainman, Karen Lastmann Assaraf, Sarai Mizrachi and Benjamin Wang 93

Deep Metric Learning to Hierarchically Rank - An Application in Product Retrieval
Kee Kiat Koo, Ashutosh Joshi, Nishaanth Reddy, Karim Bouyarmane, Ismail Tutar, Vaclav Petricek and Changhe Yuan 104

A Pretrained Language Model for Cyber Threat Intelligence
Youngja Park and Weiqiu You 113

SAMP: A Model Inference Toolkit of Post-Training Quantization for Text Processing via Self-Adaptive Mixed-Precision
Rong Tian, Zijing Zhao, Weijie Liu, Haoyan Liu, Weiquan Mao, Zhe Zhao and Kan Zhou . . . 123

KD-Boost: Boosting Real-Time Semantic Matching in E-commerce with Knowledge Distillation
Sanjay Agrawal, Vivek Sembium and Ankith M S 131

Multi-teacher Distillation for Multilingual Spelling Correction
Jingfen Zhang, Xuan Guo, Sravan Bodapati and Christopher Potts 142

Does Named Entity Recognition Truly Not Scale Up to Real-world Product Attribute Extraction?
Wei-Te Chen, Keiji Shinzato, Naoki Yoshinaga and Yandi Xia 152

Investigating Table-to-Text Generation Capabilities of Large Language Models in Real-World Information Seeking Scenarios
Yilun Zhao, Haowei Zhang, Shengyun Si, Linyong Nan, Xiangru Tang and Arman Cohan . . . 160

TMID: A Comprehensive Real-world Dataset for Trademark Infringement Detection in E-Commerce
Tongxin Hu, Zhuang Li, Xin Jin, Lizhen Qu and Xin Zhang 176

Joint Dialogue Topic Segmentation and Categorization: A Case Study on Clinical Spoken Conversations
Zhengyuan Liu, Siti Umairah Md Salleh, Hong Choon Oh, Pavitra Krishnaswamy and Nancy Chen 185

AdapterDistillation: Non-Destructive Task Composition with Knowledge Distillation
Junjie Wang, Yicheng Chen, Wangshu Zhang, Sen Hu, Teng Xu and Jing Zheng 194

PROMINET: Prototype-based Multi-View Network for Interpretable Email Response Prediction
Yuqing Wang, Prashanth Vijayaraghavan and Ehsan Degan 202

Retrieval-Enhanced Dual Encoder Training for Product Matching
Justin Chiu 216

WordArt Designer: User-Driven Artistic Typography Synthesis using Large Language Models
Jun-Yan He, Zhi-Qi Cheng, Chenyang Li, Jingdong Sun, Wangmeng Xiang, Xianhui Lin, Xiaoyang Kang, Zengke Jin, Yusen Hu, Bin Luo, Yifeng Geng and Xuansong Xie 223

Lattice Path Edit Distance: A Romanization-aware Edit Distance for Extracting Misspelling-Correction Pairs from Japanese Search Query Logs
Nobuhiro Kaji 233

Learning Multilingual Sentence Representations with Cross-lingual Consistency Regularization
Pengzhi Gao, Liwen Zhang, Zhongjun He, Hua Wu and Haifeng Wang 243

Unveiling Identity Biases in Toxicity Detection : A Game-Focused Dataset and Reactivity Analysis Approach
Josiane Van Dorpe, Zachary Yang, Nicolas Grenon-Godbout and Grégoire Winterstein 263

ORANGE: Text-video Retrieval via Watch-time-aware Heterogeneous Graph Contrastive Learning
Yucheng Lin, Tim Chang, Yaning Chang, Jianqiang Ma, Donghui Li, Ting Peng, Zang Li, Zhiyi Zhou and Feng Wang 275

Compute-Efficient Churn Reduction for Conversational Agents
Christopher Hidey and Sarthak Sarthak 284

Empower Large Language Model to Perform Better on Industrial Domain-Specific Question Answering
Fangkai Yang, Pu Zhao, Zezhong Wang, Lu Wang, Bo Qiao, Jue Zhang, Mohit Garg, Qingwei Lin, Saravan Rajmohan and Dongmei Zhang 294

Enhancing Extreme Multi-Label Text Classification: Addressing Challenges in Model, Data, and Evaluation
Dan Li, Zi Long Zhu, Janneke van de Loo, Agnes Masip Gomez, Vikrant Yadav, Georgios Tsatsaronis and Zubair Afzal 313

Query-aware Multi-modal based Ranking Relevance in Video Search
Chengcan Ye, Ting Peng, Tim Chang, Zhiyi Zhou and Feng Wang 322

Coordinated Replay Sample Selection for Continual Federated Learning
Jack Good, Jimit Majmudar, Christophe Dupuy, Jixuan Wang, Charith Peris, Clement Chung, Richard Zemel and Rahul Gupta 331

Building Real-World Meeting Summarization Systems using Large Language Models: A Practical Perspective
Md Tahmid Rahman Laskar, Xue-Yong Fu, Cheng Chen and Shashi Bhushan TN 343

Creator Context for Tweet Recommendation
Spurthi Amba Hombaiah, Tao Chen, Mingyang Zhang, Michael Bendersky, Marc Najork, Matt Colen, Sergey Levi, Vladimir Ofitserov and Tanvir Amin 353

AdaBERT-CTC: Leveraging BERT-CTC for Text-Only Domain Adaptation in ASR
Tyler Vuong, Karel Mundnich, Dhanush Bekal, Veera Raghavendra Elluru, Srikanth Ronanki and Sravan Bodapati 364

Conversing with databases: Practical Natural Language Querying
Denis Kochedykov, Fenglin Yin and Sreevidya Khatravath 372

AART: AI-Assisted Red-Teaming with Diverse Data Generation for New LLM-powered Applications
Bhaktipriya Radharapu, Kevin Robinson, Lora Aroyo and Preethi Lahoti 380

Speakerly: A Voice-based Writing Assistant for Text Composition
Dhruv Kumar, Vipul Raheja, Alice Kaiser-Schatzlein, Robyn Perry, Apurva Joshi, Justin Hugues-Nuger, Samuel Lou and Navid Chowdhury 396

Are ChatGPT and GPT-4 General-Purpose Solvers for Financial Text Analytics? A Study on Several Typical Tasks
Xianzhi Li, Samuel Chan, Xiaodan Zhu, Yulong Pei, Zhiqiang Ma, Xiaomo Liu and Sameena Shah 408

CL-QR: Cross-Lingual Enhanced Query Reformulation for Multi-lingual Conversational AI Agents
Zhongkai Sun, Zhengyang Zhao, Sixing Lu, Chengyuan Ma, Xiaohu Liu, Xing Fan, Wei Shen and Chenlei Guo 423

Improving Contextual Query Rewrite for Conversational AI Agents through User-preference Feedback Learning
Zhongkai Sun, Yingxue Zhou, Jie Hao, Xing Fan, Yanbin Lu, Chengyuan Ma, Wei Shen and Chenlei Guo 432

Scaling Neural ITN for Numbers and Temporal Expressions in Tamil: Findings for an Agglutinative Low-resource Language
Bhavuk Singhal, Sindhuja Gopalan, Amrith Krishna and Malolan Chetlur 440

EELBERT: Tiny Models through Dynamic Embeddings
Gabrielle Cohn, Rishika Agarwal, Deepanshu Gupta and Siddharth Patwardhan 451

Gold Standard Bangla OCR Dataset: An In-Depth Look at Data Preprocessing and Annotation Processes
Hasmot Ali, AKM Shahariar Azad Rabby, Md Majedul Islam, A.K.M Mahamud, Nazmul Hasan and Fuad Rahman 460

PILLOW: Enhancing Efficient Instruction Fine-tuning via Prompt Matching
Zhenting Qi, Xiaoyu Tan, Shaojie Shi, Chao Qu, Yinghui Xu and Yuan Qi 471

Welcome to the Real World: Efficient, Incremental and Scalable Key Point Analysis
Lilach Eden, Yoav Kantor, Matan Orbach, Yoav Katz, Noam Slonim and Roy Bar-Haim 483

Automatic Linking of Judgements to UK Supreme Court Hearings
Hadeel Saadany and Constantin Orasan 492

Automatic Marketing Theme and Commodity Construction System for E-commerce
Zhiping Wang, Peng Lin, Hainan Zhang, Hongshen Chen, Tianhao Li, Zhuoye Ding, Sulong Xu and Jinghe Hu 501

Towards Safer Operations: An Expert-involved Dataset of High-Pressure Gas Incidents for Preventing Future Failures
Shumpei Inoue, Minh-Tien Nguyen, Hiroki Mizokuchi, Tuan-Anh D. Nguyen, Huu-Hiep Nguyen and Dung Le 509

An Auxiliary Task Boosted Multi-task Learning Method for Service Account Retrieval with Limited Human Annotation
Yuanzhou Yao, Zhao Zhang, Kaijia Yang, Huasheng Liang, Qiang Yan and Yongjun Xu 522

VKIE: The Application of Key Information Extraction on Video Text
Siyu An, Ye Liu, Haoyuan Peng and Di Yin 532

Investigating the Role and Impact of Disfluency on Summarization
Varun Nathan, Ayush Kumar and Jithendra Vepa 541

InsightNet : Structured Insight Mining from Customer Feedback
Sandeep Sricharan Mukku, Manan Soni, Chetan Aggarwal, Jitenkumar Rana, Promod Yenigalla, Rashmi Patange and Shyam Mohan 552

E2E Spoken Entity Extraction for Virtual Agents
Karan Singla, Yeon-Jun Kim and Srinivas Bangalore 567

Generative Models for Product Attribute Extraction
Ansel Blume, Nasser Zalmout, Heng Ji and Xian Li 575

CarExpert: Leveraging Large Language Models for In-Car Conversational Question Answering
Md Rashad Al Hasan Rony, Christian Suess, Sinchana Ramakanth Bhat, Viju Sudhi, Julia Schneider, Maximilian Vogel, Roman Teucher, Ken E. Friedl and Soumya Sahoo 586

BUSTER: a BUSiness Transaction Entity Recognitiondataset
Andrea Zugarini, Andrew Zamai, Marco Ernandes and Leonardo Rigutini 605

Multi-word Tokenization for Sequence Compression
Leonidas Gee, Leonardo Rigutini, Marco Ernandes and Andrea Zugarini 612

JarviX: A LLM No code Platform for Tabular Data Analysis and Optimization
Shang-Ching Liu, ShengKun Wang, Tsungyao Chang, Wenqi Lin, Chung-Wei Hsiung, Yi-Chen Hsieh, Yu-Ping Cheng, Sian-Hong Luo and Jianwei Zhang 622

Retrieve and Copy: Scaling ASR Personalization to Large Catalogs
Sai Muralidhar Jayanthi, Devang Kulshreshtha, Saket Dingliwal, Srikanth Ronanki and Sravan Bodapati 631

STEER: Semantic Turn Extension-Expansion Recognition for Voice Assistants
Leon Zhang, Jiarui Lu, Joel Ruben Antony Moniz, Aditya Kulkarni, Dhivya Piraviperumal, Tien Dung Tran, Nick Tzou and Hong Yu 640

Self-Criticism: Aligning Large Language Models with their Understanding of Helpfulness, Honesty, and Harmlessness
Xiaoyu Tan, Shaojie Shi, Xihe Qiu, Chao Qu, Zhenting Qi, Yinghui Xu and Yuan Qi 650

InstructPTS: Instruction-Tuning LLMs for Product Title Summarization
Besnik Fetahu, Zhiyu Chen, Oleg Rokhlenko and Shervin Malmasi 663

LLM4Vis: Explainable Visualization Recommendation using ChatGPT
Lei Wang, Songheng Zhang, Yun Wang, Ee-Peng Lim and Yong Wang 675

DUBLIN: Visual Document Understanding By Language-Image Network
Kriti Aggarwal, Aditi Khandelwal, Kumar Tanmay, Owais Khan Mohammed, Qiang Liu, Monojit Choudhury, Hardik Hansrajbhai Chauhan, Subhojit Som, Vishrav Chaudhary and Saurabh Tiwary 693

DocumentNet: Bridging the Data Gap in Document Pre-training
Lijun Yu, Jin Miao, Xiaoyu Sun, Jiayi Chen, Alexander Hauptmann, Hanjun Dai and Wei Wei707

Relevance-assisted Generation for Robust Zero-shot Retrieval
Jihyuk Kim, Minsoo Kim, Joonsuk Park and Seung-won Hwang 723

Too much of product information : Don't worry, let's look for evidence!
Aryan Jain, Jitenkumar Rana and Chetan Aggarwal 732

Harnessing LLMs for Temporal Data - A Study on Explainable Financial Time Series Forecasting
Xinli Yu, Zheng Chen and Yanbin Lu 739

ViGPTQA - State-of-the-Art LLMs for Vietnamese Question Answering: System Overview, Core Models Training, and Evaluations
Minh Thuan Nguyen, Khanh Tung Tran, Nhu Van Nguyen and Xuan-Son Vu 754

An Integrated Search System for Korea Weather Data
Jinkyung Jo, Dayeon Ki, Soyoung Yoon and Minjoon Seo 765

Adaptive Hyper-parameter Learning for Deep Semantic Retrieval
Mingming Li, Chunyuan Yuan, Huimu Wang, Peng Wang, Jingwei Zhuo, Binbin Wang, Lin Liu and Sulong Xu 775

On Sample-Efficient Code Generation
Hojae Han, Yu Jin Kim, Byoungjip Kim, Youngwon Lee, Kyungjae Lee, Kyungmin Lee, Moontae Lee, Kyunghoon Bae and Seung-won Hwang 783

Batch Prompting: Efficient Inference with Large Language Model APIs
Zhoujun Cheng, Jungo Kasai and Tao Yu 792

Graph Meets LLM: A Novel Approach to Collaborative Filtering for Robust Conversational Understanding
Zheng Chen, Ziyan Jiang, Fan Yang, Eunah Cho, Xing Fan, Xiaojiang Huang, Yanbin Lu and Aram Galstyan 811

DELPHI: Data for Evaluating LLMs' Performance in Handling Controversial Issues
David Sun, Artem Abzaliev, Hadas Kotek, Christopher Klein, Zidi Xiu and Jason D Williams820

Angel: Enterprise Search System for the Non-Profit Industry
Saiful Haq, Ashutosh Sharma and Pushpak Bhattacharyya 828