

EMNLP 2022

EMNLP 2022 Industry Track

Proceedings of the Conference

December 7 - 11, 2022

©2022 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 978-1-952148-25-5

Organizing Committee

Angeliki Lazaridou, DeepMind
Yun Yao Li, Apple

Program Committee

Area Chairs

Alan Akbik, Humboldt-Universität zu Berlin
Jennifer Chu-Carroll, Elemental Cognition
Lisa Anne Hendricks, DeepMind
Marina Danilevsky, IBM Research
Mo Yu, IBM Research
Rishita Anubhai, Amazon
Ruoming Pang, Apple
Sachin Agarwal, Apple
Shimei Pan, UMBC
Srinivas Bangalore, Interactions Corp
Xiaoqiang Luo, Google

Program Committee

Abdalghani Abujabal, Amazon Alexa AI
Abeba Birhane, University College Dublin
Abhishek Singh, Samsung
Ai Ti Aw, Institute for Infocomm Research
Alex Marin, Microsoft Corporation
Alicia Sagae, Research Scientist
Amar Prakash Azad, IBM AI Research
Anastassia Loukina, Grammarly Inc
Anil Ramakrishna, Amazon
Anjishnu Kumar, Amazon Alexa
Ankur Gandhe, Amazon

Ankush Gupta, IBM Research
Anmol Goel, IIIT Hyderabad
Anna Lisa Gentile, IBM Research Almaden
Anusha Balakrishnan, Inflection AI
Aoife Cahill, Dataminr
Aparna Elangovan, The University of Melbourne
Arun Babu, Facebook
Ashish Shenoy, Meta
Aswarth Abhilash Dara, Amazon
Avneesh Saluja, Netflix
Benjamin Rozonoyer, University of Massachusetts Amherst
Bowei Zou, Institute for Infocomm Research
Brian Ulicny, Raytheon BBN Technologies
Budhaditya Deb, Microsoft Corporation
Byung-Hak Kim, AKASA
Chanjun Park, Upstage
Chen Liu, Technische Universität Darmstadt
Cheoneum Park, Hyundai Motor Group
Claudia Borg, University of Malta
Dakuo Wang, IBM Research
Damianos Karakos, Raytheon BBN Technologies
David Melville, Elemental Cognition
David Elson, Google
David Uthus, Google Research
Daxin Jiang, STCA, Microsoft
Deborah Dahl, Conversational Technologies
Deepak Muralidharan, Apple
Dejing Dou, University of Oregon
Derrick Higgins, Illinois Institute of Technology
Deyi Xiong, Tianjin University
Dongchan Kim, Amazon Alexa
Dookun Park, Amazon Alexa
Ehud Reiter, University of Aberdeen
Elio Querze, Bose Corp
Emre Barut, Alexa AI
Enrique Henestroza Anguiano, Ask Media Group
Eshwar Shamanna Girishekar, Amazon
Estevam Hruschka, Megagon Labs - <https://megagon.ai/>
Ethan Selfridge, LivePerson
Eunah Cho, Amazon, Alexa AI
Fabio Mercorio, University of Milano-Bicocca
Faisal Ladhak, Columbia University
Feifei Pan, Rensselaer Polytechnic Institute
Frank Schilder, Thomson Reuters
Frederic Mailhot, Dialpad, Inc
Geewook Kim, NAVER
Guoyin Wang, Amazon Alexa AI
Han Li, Amazon
Hassan Sawaf, aixplain, inc.
He Xie, Amazon Alexa AI
Hemant Misra, Swiggy (BundlTechnologies)

Hidetaka Kamigaito, Nara Institute of Science and Technology
Honglei Guo, Tsinghua University
ilknur Durgar Elkahlout, TURKCELL
Isabel Trancoso, INESC-ID / IST Univ. Lisbon
Ishan Jindal, IBM Research
Ismini Lourentzou, Virginia Tech
Jacopo Tagliabue, Tooso
Jade Abbott, Retro Rabbit
Jaegul Choo, KAIST
Jaesik Choi, Korea Advanced Institute of Science and Technology
Jaime Lorenzo-Trueba, Amazon
Jared Kramer, Amazon
Jesse Vig, Salesforce Research
Jiangning Chen, Amazon
Jiaze Chen, Bytedance AI Lab
Jie Ma, AWS AI Lab
John Chen, Interactions LLC
Judith Gaspers, Amazon
Jun Seok Kang, Blink Health
Justin Chiu, Rakuten USA
Kai Yu, Shanghai Jiao Tong University
Kalpa Gunaratna, Samsung Research America
Kartik Mehta, Amazon
Kasturi Bhattacharjee, AWS AI, Amazon
Kazuya Kawakami, University of Oxford
Keith Trnka, 98point6
Kfir Bar, College of Management Academic Studies
Khadige Abboud, Amazon Alexa AI
Kunho Kim, Microsoft Corporation
Laurel Orr, Stanford
Lee Becker, Educational Testing Service
Lei Chen, Rakuten
Lei Shu, Google Research
Li Dong, Amazon.com
Liangming Pan, National University of Singapore
Lin Pan, Amazon
Ling Tsou, Zoom Video Communications
Lingjia Deng, Bloomberg L.P.
Lisheng Fu, Amazon
Long Qin, Alibaba
Lorenzo Malandri, University of Milan - Bicocca
Luoxin Chen, Amazon Alexa AI
Lynette Hirschman, MITRE
Maeda Hanafi, IBM
Mahnoosh Mehrabani, Interactions LLC
Marek Suppa, Comenius University in Bratislava
Margot Mieskes, University of Applied Sciences, Darmstadt
MARK SAMMONS, Elemental Cognition
Matthew Mulholland, Educational Testing Service
Michael Flor, Educational Testing Service
Michal Shmueli-Scheuer, IBM Research

Mingyue Shang, Amazon
Mohamed Abdelhady, Amazon
Mohamed AlTantawy, Agolo
Mohammad Kachuee, Amazon Alexa AI
Monica Sunkara, Amazon
Narges Tabari, AWS AI Labs, Amazon
Navid Nobani, University of Milano-Bicocca
Ngoc Phuoc An Vo, IBM Research
Nikhil Rasiwasia, Amazon.com
Nikita Bhutani, Megagon Labs
Nitin Madnani, Educational Testing Service
Nyalleng Moorosi, Google AI
Pablo Duboue, Textualization Software Ltd.
Pengfei Liu, Centre for Perceptual and Interactive Intelligence
Pengfei Li, Nanyang Technological University
Peter Grasch, Apple Inc.
Petr Lorenc, Czech Technical University in Prague
Poornima Chozhiyath Raman, Roku
Prasanna kumar Muthukumar, BBN Technologies
Qingkai Zeng, University of Notre Dame
Radhika Gaonkar, Microsoft Search, Assistant and Intelligence
Radityo Eko Prasajo, Pitik.id
Rahul Divekar, Educational Testing Service
Ramy Eskander, Twitter
Rutuja Ubale, Educational Testing Service
Sagnik Ray Choudhury, University of Michigan
Saleh Soltan, Amazon Alexa
Saloni Potdar, IBM Watson, Carnegie Mellon University
Sandesh Swamy, Amazon
Sangameshwar Patil, TRDDC, TCS Research and Innovation
Sanjeev Kumar, Quark.ai
Sanjika Hewavitharana, eBay
Sarah Campbell, Amazon Alexa AI
Sarasi Lalithsena, IBM
Sashank Santhanam, University of North Carolina at Charlotte/ Apple
Saurabh Khanwalkar, Course Hero
Sergio Oramas, Pandora
Shashank Gupta, Microsoft Search, Assistant and Intelligence
Shuangyong Song, JD AI Research
Shuyan Dong, Meta
Siddharth Varia, Amazon
Siddharth Patwardhan, Apple
Sidharth Mudgal, Google, Inc.
Sopan Khosla, Amazon Web Services, Amazon Inc
Sourish Chaudhuri, Google Inc
Spyros Matsoukas, Amazon.com
Stavroula Skylaki, Thomson Reuters Labs
Stephen Pulman, Apple Inc.
Stevie Bergman, DeepMind
Sudarshan R. Thitte, IBM
Sumanth Prabhu, Applied Research Department, Swiggy (BUNDL Technologies), Bangalore, Karnataka,

India - 560103
Sun Kim, Naver; NCBI/NIH
Sunayana Sitaram, Microsoft Research India
Tao Wang, ByteDance AI Lab
Tara Taghavi, Amazon
Tarec Fares, Bloomberg LP
Tilman Becker, DFKI
Tirthankar Dasgupta, Tata Consultancy Services Ltd.
Tong Guo, Meituan
Tong Wang, Amazon
varun embar, Apple Inc.
Varun Kumar, Amazon Alexa
Varun Nagaraj Rao, Princeton University
Victor Soto, Amazon Inc.
Vinayshekhar Bannihatti Kumar, AWS AI
Vitobha Munigala, Research Engineer, IBM Research
Wei Zhu, ECNU
Weiwei Guo, LinkedIn
Wonseok Hwang, LBox
Xian-Ling Mao, Beijing Institute of Technology
Xiao Yang, Meta Platforms
Xiaohu Liu, Amazon
Xiliang Zhu, Dialpad
Xuan Zhu, Amazon
Xuye Liu, University of Waterloo
Yi-Chia Wang, Facebook AI
Yichao Zhou, Google Research
Yinfei Yang, Google
Yoav Katz, IBM Research AI
Youngja Park, IBM T. J. Watson Research Center
Yu Wang, Samsung Research America
Yuanfeng Song, Hong Kong University of Science and Technology, WeBank Co., Ltd
Yuji Matsumoto, Riken Center for Advanced Intelligence Project
Yuval Marton, University of Washington
Ziming Huang, Tencent

Table of Contents

<i>Unsupervised Term Extraction for Highly Technical Domains</i> Francesco Fusco, Peter Staar and Diego Antognini	1
<i>DynaMaR: Dynamic Prompt with Mask Token Representation</i> Xiaodi Sun, Sunny Rajagopalan, Priyanka Nigam, Weiyi Lu, Yi Xu, Iman Keivanloo, Belinda Zeng and Trishul Chilimbi	9
<i>A Hybrid Approach to Cross-lingual Product Review Summarization</i> Saleh Soltan, Victor Soto, Ke Tran and Wael Hamza	18
<i>Augmenting Operations Research with Auto-Formulation of Optimization Models From Problem Descriptions</i> Rindra Ramamonjison, Haley Li, Timothy Yu, Shiqi HE, Vishnu Rengan, Amin Banitalebi-Dehkordi, Zirui Zhou and Yong Zhang	29
<i>Knowledge Distillation based Contextual Relevance Matching for E-commerce Product Search</i> Ziyang Liu, Chaokun Wang, Hao Feng, Lingfei Wu and Liqun Yang	63
<i>Accelerating the Discovery of Semantic Associations from Medical Literature: Mining Relations Between Diseases and Symptoms</i> Alberto Purpura, Francesca Bonin and Joao Bettencourt-Silva	77
<i>PENTATRON: PErsonalized coNText-Aware Transformer for Retrieval-based cOnversational uNderstanding</i> Niranjan Uma Naresh, Ziyang Jiang, Ankit Ankit, Sungjin Lee, Jie Hao, Xing Fan and Chenlei Guo	90
<i>Machine translation impact in E-commerce multilingual search</i> Bryan Zhang and amita misra	99
<i>Ask-and-Verify: Span Candidate Generation and Verification for Attribute Value Extraction</i> Yifan Ding, Yan Liang, Nasser Zalmout, Xian Li, Christan Grant and Tim Wenginger	110
<i>Consultation Checklists: Standardising the Human Evaluation of Medical Note Generation</i> Aleksandar Savkov, Francesco Moramarco, Alex Papadopoulos Korfiatis, Mark Perera, Anya Belz and Ehud Reiter	121
<i>Towards Need-Based Spoken Language Understanding Model Updates: What Have We Learned?</i> Quynh Do, Judith Gaspers, Daniil Sorokin and Patrick Lehnen	131
<i>Knowledge Distillation Transfer Sets and their Impact on Downstream NLU Tasks</i> Charith Peris, Lizhen Tan, Thomas Gueudre, Turan Gojayev, Pan Wei and Gokmen Oz	138
<i>Exploiting In-Domain Bilingual Corpora for Zero-Shot Transfer Learning in NLU of Intra-Sentential Code-Switching Chatbot Interactions</i> Maia Aguirre, Manex Serras, Laura García-Sardiña, Jacobo López-Fernández, Ariane Méndez and Arantza del Pozo	148
<i>Calibrating Imbalanced Classifiers with Focal Loss: An Empirical Study</i> Cheng Wang, Jorge Balazs, György Szarvas, Patrick Ernst, Lahari Poddar and Pavel Danchenko	155

<i>Unsupervised training data re-weighting for natural language understanding with local distribution approximation</i>	
Jose Garrido Ramas, Dieu-Thu Le, Bei Chen, Manoj Kumar and Kay Rottmann	164
<i>Cross-Encoder Data Annotation for Bi-Encoder Based Product Matching</i>	
Justin Chiu and Keiji Shinzato	171
<i>Deploying a Retrieval based Response Model for Task Oriented Dialogues</i>	
Lahari Poddar, György Szarvas, Cheng Wang, Jorge Balazs, Pavel Danchenko and Patrick Ernst	179
<i>Tackling Temporal Questions in Natural Language Interface to Databases</i>	
Ngoc Phuoc An Vo, Octavian Popescu, Irene Manotas and Vadim Sheinin	189
<i>Multi-Tenant Optimization For Few-Shot Task-Oriented FAQ Retrieval</i>	
Asha Vishwanathan, Rajeev Warriar, Gautham Vadakkekara Suresh and Chandra Shekhar Kandpal	198
<i>Iterative Stratified Testing and Measurement for Automated Model Updates</i>	
Elizabeth Dekeyser, Nicholas Comment, Shermin Pei, Rajat Kumar, Shruti Rai, Fengtao Wu, Lisa Haverty and Kanna Shimizu	208
<i>SLATE: A Sequence Labeling Approach for Task Extraction from Free-form Inked Content</i>	
Apurva Gandhi, Ryan Serrao, Biyi Fang, Gilbert Antonius, Jenna Hong, Tra My Nguyen, Sheng Yi, Ehi Nosakhare, Irene Shaffer, Soundararajan Srinivasan and Vivek Gupta	216
<i>Gaining Insights into Unrecognized User Utterances in Task-Oriented Dialog Systems</i>	
Ella Rabinovich, Matan Vetzler, David Boaz, Vineet Kumar, Gaurav Pandey and Ateret Anaby Tavor	228
<i>CoCoID: Learning Contrastive Representations and Compact Clusters for Semi-Supervised Intent Discovery</i>	
Qian Cao, Deyi Xiong, Qinlong Wang and Xia Peng	236
<i>Tractable & Coherent Multi-Document Summarization: Discrete Optimization of Multiple Neural Modeling Streams via Integer Linear Programming</i>	
Litton J Kurisinkel and Nancy Chen	247
<i>Grafting Pre-trained Models for Multimodal Headline Generation</i>	
Lingfeng Qiao, Chen Wu, Ye Liu, haoyuan peng, di yin and Bo Ren	254
<i>Semi-supervised Adversarial Text Generation based on Seq2Seq models</i>	
Hieu Le, Dieu-Thu Le, Verena Weber, Chris Church, Kay Rottmann, Melanie Bradford and Peter Chin	264
<i>Is it out yet? Automatic Future Product Releases Extraction from Web Data</i>	
Gilad Fuchs, Ido Ben-Shaul and Matan Mandelbrod	273
<i>Automatic Scene-based Topic Channel Construction System for E-Commerce</i>	
Peng Lin, Yanyan Zou, Lingfei Wu, Mian Ma, Zhuoye Ding and Bo Long	282
<i>SpeechNet: Weakly Supervised, End-to-End Speech Recognition at Industrial Scale</i>	
Raphael Tang, Karun Kumar, Gefei Yang, Akshat Pandey, Yajie Mao, Vladislav Belyaev, Madhuri Emmadi, Craig Murray, Ferhan Ture and Jimmy Lin	295

<i>Controlled Language Generation for Language Learning Items</i> Kevin Stowe, Debanjan Ghosh and Mengxuan Zhao	304
<i>Improving Text-to-SQL Semantic Parsing with Fine-grained Query Understanding</i> Jun Wang, Patrick Ng, Alexander Hanbo Li, Jiarong Jiang, Zhiguo Wang, Bing Xiang, Ramesh Nallapati and Sudipta Sengupta	316
<i>Unsupervised Dense Retrieval for Scientific Articles</i> Dan Li, Vikrant Yadav, Zubair Afzal and George Tsatsaronis	323
<i>Learning Geolocations for Cold-Start and Hard-to-Resolve Addresses via Deep Metric Learning</i> Govind . and Saurabh Sohoney	332
<i>Meta-learning Pathologies from Radiology Reports using Variance Aware Prototypical Networks</i> Arijit Sehanobish, Kawshik Kannan, Nabila Abraham, Anasuya Das and Benjamin Odry	342
<i>Named Entity Recognition in Industrial Tables using Tabular Language Models</i> Aneta Koleva, Martin Ringsquandl, Mark Buckley, Rakeb Hasan and Volker Tresp	358
<i>Reinforced Question Rewriting for Conversational Question Answering</i> Zhiyu Chen, Jie Zhao, Anjie Fang, Besnik Fetahu, Oleg Rokhlenko and Shervin Malmasi	367
<i>Improving Large-Scale Conversational Assistants using Model Interpretation based Training Sample Selection</i> Stefan Schroedl, Manoj Kumar, Kiana Hajebi, Morteza Ziyadi, Sriram Venkatapathy, Anil Ramakrishna, Rahul Gupta and Pradeep Natarajan	381
<i>Improving Precancerous Case Characterization via Transformer-based Ensemble Learning</i> Yizhen Zhong, Jiajie Xiao, Thomas Vetterli, Mahan Matin, Ellen Loo, Jimmy Lin, Richard Bourgon and Ofer Shapira	389
<i>Developing Prefix-Tuning Models for Hierarchical Text Classification</i> Lei Chen, Houwei Chou and Xiaodan Zhu	400
<i>PAIGE: Personalized Adaptive Interactions Graph Encoder for Query Rewriting in Dialogue Systems</i> Daniel Biś, Saurabh Gupta, Jie Hao, Xing Fan and Chenlei Guo	408
<i>Fast Vocabulary Transfer for Language Model Compression</i> Leonidas Gee, Andrea Zugarini, Leonardo Rigutini and Paolo Torrioni	419
<i>Multimodal Context Carryover</i> Prashan Wanigasekara, Nalin Gupta, Fan Yang, Emre Barut, Zeynab Raeesy, Kechen Qin, Stephen Rawls, Xinyue Liu, Chengwei Su and Spurthi Sandiri	427
<i>Distilling Multilingual Transformers into CNNs for Scalable Intent Classification</i> Besnik Fetahu, Akash Veeragouni, Oleg Rokhlenko and Shervin Malmasi	439
<i>Bringing the State-of-the-Art to Customers: A Neural Agent Assistant Framework for Customer Service Support</i> Stephen Obadinma, Faiza Khan Khattak, Shirley Wang, Tania Sidhorn, Elaine Lau, Sean Robertson, Jingcheng Niu, Winnie Au, Alif Munim, Karthik Raja Kalaiselvi Bhaskar, Bencheng Wei, Iris Ren, Muhammad Waqar, Erin Li, Bukola Ishola, Michael Wang, Griffin Tanner, Yu-Jia Shiah, Sean X. Zhang, Kwesi Apponsah, Kanishk Patel, Jaswinder Narain, Pandya Deval, Xiaodan Zhu, Frank Rudzicz and Elham Dolatabadi	450

<i>Zero-Shot Dynamic Quantization for Transformer Inference</i> Yousef El-Kurdi, Jerry Quinn and Avi Sil	461
<i>Fact Checking Machine Generated Text with Dependency Trees</i> Alex Estes, Nikhita Vedula, Marcus Collins, Matt Cecil and Oleg Rokhlenko	468
<i>Prototype-Representations for Training Data Filtering in Weakly-Supervised Information Extraction</i> Nasser Zalmout and Xian Li	477
<i>CGF: Constrained Generation Framework for Query Rewriting in Conversational AI</i> Jie Hao, Yang Liu, Xing Fan, Saurabh Gupta, Saleh Soltan, Rakesh Chada, Pradeep Natarajan, Chenlei Guo and Gokhan Tur	485
<i>Entity-level Sentiment Analysis in Contact Center Telephone Conversations</i> Xue-Yong Fu, Cheng Chen, Md Tahmid Rahman Laskar, Shayna Gardiner, Pooja Hiranandani and Shashi Bhushan TN	494
<i>QUILL: Query Intent with Large Language Models using Retrieval Augmentation and Multi-stage Distillation</i> Krishna Srinivasan, Karthik Raman, Anupam Samanta, Lingrui Liao, Luca Bertelli and Michael Bendersky	502
<i>Distinguish Sense from Nonsense: Out-of-Scope Detection for Virtual Assistants</i> Cheng Qian, Haode Qi, Gengyu Wang, Ladislav Kunc and Saloni Potdar	512
<i>PLATO-Ad: A Unified Advertisement Text Generation Framework with Multi-Task Prompt Learning</i> Zeyang Lei, Chao Zhang, Xinchao Xu, Wenquan Wu, Zheng-Yu Niu, Hua Wu, Haifeng Wang, Yi Yang and Shuanglong Li	522
<i>Dense Feature Memory Augmented Transformers for COVID-19 Vaccination Search Classification</i> Jai Gupta, Yi Tay, Chaitanya Kamath, Vinh Tran, Donald Metzler, Shailesh Bavadekar, Mimi Sun and Evgeniy Gabrilovich	531
<i>Full-Stack Information Extraction System for Cybersecurity Intelligence</i> Youngja Park and Taesung Lee	541
<i>Deploying Unified BERT Moderation Model for E-Commerce Reviews</i> Ravindra Nayak and Nikesh Garera	550
<i>SimANS: Simple Ambiguous Negatives Sampling for Dense Text Retrieval</i> Kun Zhou, Yeyun Gong, Xiao Liu, Wayne Xin Zhao, Yelong Shen, Anlei Dong, Jingwen Lu, Rangan Majumder, Ji-Rong Wen, Nan Duan and Weizhu Chen	558
<i>Revisiting and Advancing Chinese Natural Language Understanding with Accelerated Heterogeneous Knowledge Pre-training</i> Taolin Zhang, junwei dong, Jianing Wang, Chengyu Wang, Ang Wang, Yinghui Liu, jun huang, Yong Li and XIAOFENG HE	570
<i>A Stacking-based Efficient Method for Toxic Language Detection on Live Streaming Chat</i> Yuto Oikawa, Yuki Nakayama and Koji Murakami	581
<i>End-to-End Speech to Intent Prediction to improve E-commerce Customer Support Voicebot in Hindi and English</i> Abhinav Goyal, Anupam Singh and Nikesh Garera	589

<i>PILE: Pairwise Iterative Logits Ensemble for Multi-Teacher Labeled Distillation</i>	
Lianshang Cai, Linhao Zhang, Dehong Ma, Jun Fan, Daiting Shi, Yi Wu, Zhicong Cheng, Simiu Gu and Dawei Yin	597
<i>A Comprehensive Evaluation of Biomedical Entity-centric Search</i>	
Elena Tutubalina, Zulfat Miftahutdinov, Vladimir Muravlev and Anastasia Shneyderman	606
<i>Domain Adaptation of Machine Translation with Crowdworkers</i>	
Makoto Morishita, Jun Suzuki and Masaaki Nagata	616
<i>Biomedical NER for the Enterprise with Distillated BERN2 and the Kazu Framework</i>	
Wonjin Yoon, Richard Jackson, Elliot Ford, Vladimir Poroshin and Jaewoo Kang	629
<i>Large-scale Machine Translation for Indian Languages in E-commerce under Low Resource Constraints</i>	
Amey Patil and Nikesh Garera	637
<i>Topic Modeling by Clustering Language Model Embeddings: Human Validation on an Industry Dataset</i>	
Anton Eklund and Mona Forsman	645

