

ClinicalNLP 2023

**The 5th Workshop on Clinical Natural Language Processing
(ClinicalNLP)**

Proceedings of the Workshop

July 14, 2023

The ClinicalNLP organizers gratefully acknowledge the support from the following sponsors.



©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 978-1-959429-88-3

Preface

This volume contains papers from the 5th Workshop on Clinical Natural Language Processing (Clinical NLP), held at ACL 2023.

Clinical text offers unique challenges that differentiate it not only from open-domain data, but from other types of text in the biomedical domain as well. Notably, clinical text contains a significant number of abbreviations, medical terms, and other clinical jargon. Clinical narratives are characterized by non-standard document structures that are often critical to overall understanding. Narrative provider notes are designed to communicate with other experts while at the same time serving as a legal record. Finally, clinical notes contain sensitive patient-specific information that raise privacy and security concerns that present special challenges for natural language systems. This workshop focuses on the work that develops methods to address the above challenges, with the goal of advancing state-of-the-art in clinical NLP.

ClinicalNLP 2023 also hosted the MEDIQA-Chat 2023 shared tasks that promote research on effective solutions for clinical note generation from medical conversations. The shared tasks focused on the summarization of doctor-patient conversations and on the generation of synthetic dialogues from clinical notes for data augmentation. They introduced new benchmarks for training and evaluation and used an ensemble of evaluation metrics that highly correlate with human judgments. Further, the organizers added a new requirement to submit the code for a second evaluation of the outputs. The MEDIQA-Chat shared tasks attracted 120 registered teams with 17 teams submitting their codes and runs for official participation. The participating teams experimented with the recently released Large Language Models (LLMs) vs. older models and explored data augmentation, fine-tuning, and prompting methods. The results provided new insights on the best approaches and techniques for future research directions in clinical text generation.

This year, we received the total of 82 submissions, inclusive of shared task submissions, from which 58 were accepted for presentation.

Organizing Committee

General Chairs

Tristan Naumann, Microsoft Research
Asma Ben Abacha, Microsoft
Steven Bethard, University of Arizona
Kirk Roberts, UTHealth Houston
Anna Rumshisky, UMass Lowell

MEDIAQA-Chat 2023 Shared Task Organizers

Asma Ben Abacha, Microsoft
Wen-wai Yim, Microsoft
Griffin Adams, Columbia University
Neal Snider, Microsoft/Nuance
Meliha Yetisgen, University of Washington

Program Committee

Program Committee

Griffin Thomas Adams, Surabhi Adhikari, Ashwag Alasmari, Amal Abdullah Alqahtani, Emily Alsentzer, Ibtihel Amara, Hajer Ayadi

Steven Bethard, Sjaak Brinkkemper

Leonardo Campillos-Llanos, Salim Chemlal, Qingyu Chen, Shan Chen, Cheryl Clark

Hong-Jie Dai, Judith W Dexheimer, Dmitriy Dligach, Richard Dufour, Jocelyn Dunstan

Naome A Etori

Matúš Falis, Yadan Fan, Xavier Fontaine

Zelalem Gero, John Michael Giorgi, Natalia Grabar

Mei-Hua Hall, Sadid A. Hasan, Rim Helaoui, Ming Huang

Raphael Iyamu

Xiaoqian Jiang, Qiao Jin, Alistair Johnson

Yoshinobu Kano, Raghav Kapoor, Yejin Kim, Martin Krallinger, Shiba Kuanar

Yanis Labrak, Egoitz Laparra, Alberto Lavelli, Ulf Leser, Qiu hao Lu, Yuxing Lu

Diwakar Mahajan, Pranita Yogesh Mahajan, Gaetano Manzo, Sérgio Matos, George Michalopoulos, Kirill Milintsevich, Timothy A Miller, Ishani Mondal

Aakanksha Naik, Tristan Naumann, Aurélie Névéol

Kadir Bulut Ozler

Ankur Padia, Yifan Peng, Joseph M Plasek

Satyajeet Raje, Pavithra Rajendran, Giridhar Kaushik Ramachandran, Anku Rani, Fabio Rinaldi, Kirk Roberts, Matías Rojas, Frank Rudzicz

Guergana K Savova, Ashwyn Sharma, Atul Singh, Sonish Sivarajkumar, Sarvesh Soni, Arvind Krishna Sridhar, Dhananjay Srivastava, Karl Eric Swanson

Behrad Taghibeyglou, Neset Tan, Xiangru Tang, Khushboo Thaker, Augustin Toma

Byron C Wallace, Jinge Wu, Susmitha Wunnava

Dongfang Xu

Meliha Yetisgen, Wen-wai Yim, WonJin Yoon, Paul Youssef, Xiang Yue

Boya Zhang, Weipeng Zhou

Keynote Talk: Patient record summarization: tasks, approaches, evaluation, and open challenges

Noémie Elhadad
Columbia University

Abstract: The patient record contains an overwhelming large amount of information, too much for a clinician to make sense of it, and yet the information it contains may be critical for clinicians to care for their patients safely and effectively. In this talk, I will review two tasks to alleviate the information overload in clinical care: longitudinal patient record summarization and abstractive brief hospital course summarization. I will describe potential approaches, evaluation objectives, and current open questions. Finally, using the abstractive task of brief hospital course summarization as a grounding example, I will discuss large language models (LLMs) in the context of clinical NLP.

Bio: Noémie Elhadad is Chair of the department of Biomedical Informatics at Columbia University, affiliated with the department of Computer Science and the Data Science Institute. Elhadad's research lies at the intersection of artificial intelligence, human-centered computing, and medicine. She creates novel methods and tools to support patients and clinicians in their information needs, with particular focus on ensuring that the AI systems of the future are robust, safe, fair, and just.

Keynote Talk: The evolution of representations for clinical text and a few more thoughts about generative clinical models

Timothy Miller

Boston Children's Hospital, Harvard Medical School

Abstract: Large language models (LLMs) have excited the broader public like no previous NLP advance. This has led to predictions from all corners about the future of LLM-enabled NLP for clinical data and tasks. In this talk, I review several recent projects from my lab that did not use LLMs, and re-imagine these projects in an LLM-enabled context. The talk then synthesizes the lessons from those projects to propose some guidelines for optimal use of LLMs in clinical NLP research, imagine future directions that are now enabled, and to make some predictions about the future of our field.

Bio: Tim Miller is an Associate Professor in the Computational Health Informatics Program at Boston Children's Hospital, Department of Pediatrics at Harvard Medical School, and at the Harvard-MIT Center for Regulatory Science. He is the PI of the Machine Learning for Medical Language Lab, home of several federally funded projects, including projects focused on basic biomedical NLP research, as well as projects that are driven by biomedical use cases. His research focuses on domain adaptation/generalizability of ML-based NLP methods, as well as methods for learning universal patient representations.

Table of Contents

<i>Clinical BERTScore: An Improved Measure of Automatic Speech Recognition Performance in Clinical Settings</i> Joel Shor, Ruyue Agnes Bi, Subhashini Venugopalan, Steven Ibara, Roman Goldenberg and Ehud Rivlin	1
<i>Medical Visual Textual Entailment for Numerical Understanding of Vision-and-Language Models</i> Hitomi Yanaka, Yuta Nakamura, Yuki Chida and Tomoya Kurosawa	8
<i>Privacy-Preserving Knowledge Transfer through Partial Parameter Sharing</i> Paul Youssef, Jörg Schlötterer and Christin Seifert	19
<i>Breaking Barriers: Exploring the Diagnostic Potential of Speech Narratives in Hindi for Alzheimer’s Disease</i> Kritesh Rauniyar, Shuvam Shiwakoti, Sweta Poudel, Surendrabikram Thapa, Usman Naseem and Mehwish Nasim	24
<i>Investigating Massive Multilingual Pre-Trained Machine Translation Models for Clinical Domain via Transfer Learning</i> Lifeng Han, Gleb Erofeev , Irina Sorokina, Serge Gladkoff and Goran Nenadic	31
<i>Tracking the Evolution of Covid-19 Symptoms through Clinical Conversations</i> Ticiana Linhares Coelho Da Silva, José A. Fernandes De Macêdo and Régis Pires Magalhães .	41
<i>Aligning Factual Consistency for Clinical Studies Summarization through Reinforcement Learning</i> Xiangru Tang, Arman Cohan and Mark Gerstein	48
<i>Navigating Data Scarcity: Pretraining for Medical Utterance Classification</i> Do June Min, Veronica Perez-Rosas and Rada Mihalcea	59
<i>Hindi Chatbot for Supporting Maternal and Child Health Related Queries in Rural India</i> Ritwik Mishra, Simranjeet Singh, Jasmeet Kaur, Pushpendra Singh and Rajiv Ratn Shah	69
<i>Multi-Task Training with In-Domain Language Models for Diagnostic Reasoning</i> Brihat Sharma, Yanjun Gao, Timothy A Miller, Matthew Churpek, Majid Afshar and Dmitriy Dligach	78
<i>Context-aware Medication Event Extraction from Unstructured Text</i> Noushin Salek Faramarzi, Meet Patel, Sai Harika Bandarupally and Ritwik Banerjee	86
<i>Improving Automatic KCD Coding: Introducing the KoDAK and an Optimized Tokenization Method for Korean Clinical Documents</i> Geunyeong Jeong, Juoh Sun, Seokwon Jeong, Hyunjin Shin and Harksoo Kim	96
<i>Who needs context? Classical techniques for Alzheimer’s disease detection</i> Behrad Taghibeyglou and Frank Rudzicz	102
<i>Knowledge Injection for Disease Names in Logical Inference between Japanese Clinical Texts</i> Natsuki Murakami, Mana Ishida, Yuta Takahashi, Hitomi Yanaka and Daisuke Bekki	108
<i>Training Models on Oversampled Data and a Novel Multi-class Annotation Scheme for Dementia Detection</i> Nadine Abdelhalim, Ingy Yasser Hassan Abdou Abdelhalim and Riza Batista-Navarro	118

<i>Improving the Transferability of Clinical Note Section Classification Models with BERT and Large Language Model Ensembles</i>	
Weipeng Zhou, Majid Afshar, Dmitriy Dligach, Yanjun Gao and Timothy A Miller	125
<i>Can Large Language Models Safely Address Patient Questions Following Cataract Surgery?</i>	
Mohita Chowdhury, Ernest Lim, Aisling Higham, Rory McKinnon, Nikoletta Ventoura, Yajie Vera He and Nick De Pennington	131
<i>Large Scale Sequence-to-Sequence Models for Clinical Note Generation from Patient-Doctor Conversations</i>	
Gagandeep Singh, Yue Pan, Jesus Andres-Ferrer, Miguel Del-Agua, Frank Diehl, Joel Pinto and Paul Vozila	138
<i>clulab at MEDIQA-Chat 2023: Summarization and classification of medical dialogues</i>	
Kadir Bulut Ozler and Steven Bethard	144
<i>Leveraging Natural Language Processing and Clinical Notes for Dementia Detection</i>	
Ming Liu, Richard Beare, Taya Collyer, Nadine Andrew and Velandai Srikanth	150
<i>Automated Orthodontic Diagnosis from a Summary of Medical Findings</i>	
Takumi Ohtsuka, Tomoyuki Kajiwara, Chihiro Tanikawa, Yuujin Shimizu, Hajime Nagahara and Takashi Ninomiya	156
<i>Harnessing the Power of BERT in the Turkish Clinical Domain: Pretraining Approaches for Limited Data Scenarios</i>	
Hazal Türkmen, Oguz Dikenelli, Cenk Eraslan, Mehmet Calli and Suha Sureyya Ozbek	161
<i>A Meta-dataset of German Medical Corpora: Harmonization of Annotations and Cross-corpus NER Evaluation</i>	
Ignacio Llorca, Florian Borchert and Matthieu-P. Schapranow	171
<i>Uncovering the Potential for a Weakly Supervised End-to-End Model in Recognising Speech from Patient with Post-Stroke Aphasia</i>	
Giulia Sanguedolce, Patrick Naylor and Fatemeh Geranmayeh	182
<i>Textual Entailment for Temporal Dependency Graph Parsing</i>	
Jiarui Yao, Steven Bethard, Kristin Wright-Bettner, Eli T Goldner, David A Harris and Guergana K Savova	191
<i>Generating medically-accurate summaries of patient-provider dialogue: A multi-stage approach using large language models</i>	
Varun Nair, Elliot Schumacher and Anitha Kannan	200
<i>Factors Affecting the Performance of Automated Speaker Verification in Alzheimer’s Disease Clinical Trials</i>	
Malikeh Ehghaghi, Marija Stanojevic, Ali Akram and Jekaterina Novikova	218
<i>Team Cadence at MEDIQA-Chat 2023: Generating, augmenting and summarizing clinical dialogue with large language models</i>	
Ashwyn Sharma, David Ian Feldman and Aneesh Jain	228
<i>Method for Designing Semantic Annotation of Sepsis Signs in Clinical Text</i>	
Melissa Y. Yan, Lise Tuset Gustad, Lise Husby Høvik and Øystein Nytrø	236
<i>Prompt Discriminative Language Models for Domain Adaptation</i>	
Keming Lu, Peter Potash, Xihui Lin, Yuwen Sun, Zihan Qian, Zheng Yuan, Tristan Naumann, Tianxi Cai and Junwei Lu	247

<i>Cross-domain German Medical Named Entity Recognition using a Pre-Trained Language Model and Unified Medical Semantic Types</i>	
Siting Liang, Mareike Hartmann and Daniel Sonntag	259
<i>Reducing Knowledge Noise for Improved Semantic Analysis in Biomedical Natural Language Processing Applications</i>	
Usman Naseem, Surendrabikram Thapa, Qi Zhang, Liang Hu, Anum Masood and Mehwish Nasim	272
<i>Medical knowledge-enhanced prompt learning for diagnosis classification from clinical text</i>	
Yuxing Lu, Xukai Zhao and Jinzhuo Wang	278
<i>Multilingual Clinical NER: Translation or Cross-lingual Transfer?</i>	
Félix Gaschi, Xavier Fontaine, Parisa Rastin and Yannick Toussaint	289
<i>UMLS-KGI-BERT: Data-Centric Knowledge Integration in Transformers for Biomedical Entity Recognition</i>	
Aidan Mannion, Didier Schwab and Lorraine Goeuriot	312
<i>WangLab at MEDIQA-Chat 2023: Clinical Note Generation from Doctor-Patient Conversations using Large Language Models</i>	
John Michael Giorgi, Augustin Toma, Ronald Xie, Sondra Chen, Kevin R An, Grace Xiaoyu Zheng and BO Wang	323
<i>Automatic Coding at Scale: Design and Deployment of a Nationwide System for Normalizing Referrals in the Chilean Public Healthcare System</i>	
Fabián Villena, Matías Rojas, Felipe Arias, Jorge Pacheco, Paulina Vera and Jocelyn Dunstan	335
<i>Building blocks for complex tasks: Robust generative event extraction for radiology reports under domain shifts</i>	
Sitong Zhou, Meliha Yetisgen and Mari Ostendorf	344
<i>Intersectionality and Testimonial Injustice in Medical Records</i>	
Kenya S. Andrews, Bhuvni Shah and Lu Cheng	358
<i>Interactive Span Recommendation for Biomedical Text</i>	
Louis Blankemeier, Theodore Zhao, Robert Tinn, Sid Kiblawi, Yu Gu, Akshay S Chaudhari, Hoifung Poon, Sheng Zhang, Mu Wei and J. Samuel Preston	373
<i>Prompt-based Extraction of Social Determinants of Health Using Few-shot Learning</i>	
Giridhar Kaushik Ramachandran, Yajuan Fu, Bin Han, Kevin Lybarger, Nic J Dobbins, Ozlem Uzuner and Meliha Yetisgen	385
<i>Teddysum at MEDIQA-Chat 2023: an analysis of fine-tuning strategy for long dialog summarization</i>	
Yongbin Jeong, Ju-Hyuck Han, Kyung Min Chae, Yousang Cho, Hyunbin Seo, KyungTae Lim, Key-Sun Choi and Younggyun Hahm	394
<i>Rare Codes Count: Mining Inter-code Relations for Long-tail Clinical Text Classification</i>	
Jiamin Chen, Xuhong Li, Junting Xi, Lei Yu and Haoyi Xiong	403
<i>NewAgeHealthWarriors at MEDIQA-Chat 2023 Task A: Summarizing Short Medical Conversation with Transformers</i>	
Prakhar Mishra and Ravi Theja Desetty	414
<i>Storyline-Centric Detection of Aphasia and Dysarthria in Stroke Patient Transcripts</i>	
Peiqi Sui, Kelvin Wong, Xiaohui Yu, John Julius Volpi and Stephen T. C. Wong	422

<i>Pre-trained language models in Spanish for health insurance coverage</i> Claudio Aracena, Nicolás Rodríguez, Victor Rocco and Jocelyn Dunstan	433
<i>Utterance Classification with Logical Neural Network: Explainable AI for Mental Disorder Diagnosis</i> Yeldar Toleubay, Don Joven Agravante, Daiki Kimura, Baihan Lin, Djallel Bouneffouf and Michiaki Tatsubori	439
<i>A Survey of Evaluation Methods of Generated Medical Textual Reports</i> Yongxin Zhou, Fabien Ringeval and François Portet	447
<i>UMASS_BioNLP at MEDIQA-Chat 2023: Can LLMs generate high-quality synthetic note-oriented doctor-patient conversations?</i> Junda Wang, Zonghai Yao, Avijit Mitra, Samuel Osebe, Zhichao Yang and Hong yu	460
<i>HealthMavericks@MEDIQA-Chat 2023: Benchmarking different Transformer based models for Clinical Dialogue Summarization</i> Kunal Suri, Saumajit Saha and Atul Singh	472
<i>SummQA at MEDIQA-Chat 2023: In-Context Learning with GPT-4 for Medical Summarization</i> Yash Mathur, Sanketh Rangreji, Raghav Kapoor, Medha Palavalli, Amanda Bertsch and Matthew R. Gormley	490
<i>Overview of the MEDIQA-Chat 2023 Shared Tasks on the Summarization & Generation of Doctor-Patient Conversations</i> Asma Ben Abacha, Wen-wai Yim, Griffin Thomas Adams, Neal Snider and Meliha Yetisgen	503
<i>Transfer Learning for Low-Resource Clinical Named Entity Recognition</i> Nevasini Sasikumar and Krishna Sri Ipsit Mantri	514
<i>IUTEAM1 at MEDIQA-Chat 2023: Is simple fine tuning effective for multi layer summarization of clinical conversations?</i> Dhananjay Srivastava	519
<i>Care4Lang at MEDIQA-Chat 2023: Fine-tuning Language Models for Classifying and Summarizing Clinical Dialogues</i> Amal Abdullah Alqahtani, Rana Salama, Mona T. Diab and Abdou Youssef	524
<i>Calvados at MEDIQA-Chat 2023: Improving Clinical Note Generation with Multi-Task Instruction Finetuning</i> Kirill Milintsevich and Navneet Agarwal	529
<i>DS4DH at MEDIQA-Chat 2023: Leveraging SVM and GPT-3 Prompt Engineering for Medical Dialogue Classification and Summarization</i> Boya Zhang, Rahul Mishra and Douglas Teodoro	536
<i>GersteinLab at MEDIQA-Chat 2023: Clinical Note Summarization from Doctor-Patient Conversations through Fine-tuning and In-context Learning</i> Xiangru Tang, Andrew Tran, Jeffrey Tan and Mark Gerstein	546