BioNLP 2024

The 23rd Meeting of the ACL Special Interest Group on Biomedical Natural Language Processing

Proceedings of the Workshop and Shared Tasks

August 16, 2024

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 979-8-89176-130-8

Biomedical natural language processing in 2024: The year of BioMedGen

Dina Demner-Fushman, Sophia Ananiadou, Makoto Miwa, Kirk Roberts and Jun-ichi Tsujii

The development of Large Language Models (LLMs) applied to complex Biomedical Language Processing tasks keeps growing steadily. This growth instigated the anticipation of major breakthroughs in language generation and downstream healthcare tasks, as well as concerns with respect to potential harms and irresponsible use of AI applications. Both the medical informatics communities and regulatory agencies are developing guidelines and checklists for conducting trustworthy LLM-based research and reporting the results of this research [1] [2] [3].

The submissions to the BioNLP 2024 workshop and the Shared Tasks demonstrated once again that the workshop sponsored by the ACL Special Interest Group on Biomedical Natural Language Processing (SIGBIOMED) is the preferred venue for the groundbreaking research and applications in Biomedical Language Processing. BioNLP remains the flagship and the generalist in biomedical language processing, accepting all noteworthy work independently of the tasks and languages studied. The quality of submissions continues to impress the program committee and the organizers.

BioNLP 2024 received 61 submissions, of which six were accepted for oral presentation and 37 as poster presentations. The presentations cover a wide range of the foundational biomedical language processing research and clinical applications, exploring generation of a variety of clinical reports, extraction of information from the literature and social media, prediction of patients' outcomes and generation of datasets and benchmarks for question answering.

The Shared Tasks included generation of radiology reports (RRG24, 8 participating teams), generation of hospital course summaries and discharge instructions (Discharge Me!, 12 participating teams), and abstractive summarization of biomedical articles (BioLaySumm, 14 participating teams). The overviews of the tasks and short presentations of the best performing approaches are included in the workshop program. The participants in all Shared Tasks present their work in a dedicated poster session.

The keynote by Titipat Achakulvisut, Department of Biomedical Engineering, Mahidol University, Thailand is titled Enhancing Neuroscience Conferences through Natural Language Processing

This talk presents the development and implementation of natural language processing (NLP) tools at neuroscience conferences. Dr. Achakulvisut has successfully integrated these tools into various conferences, including a recommendation engine at the Society for Neuroscience (SfN) meeting, one-on-one matching at the Conference on Cognitive Computational Neuroscience (CCN), paper-reviewer matching for the Computational and Systems Neuroscience (COSYNE) conference, and reviewer recommendations for NBDT journal. His group employs a fine-tuning and contrastive learning approach to adapt transformer-based models, such as MiREAD and SciBERT for neuroscience. The models were evaluated using both distance metrics and recommendation arena assessments. The goal of exploring NLP tools in non-computer science domains is to enhance the interactions of researchers and attendees.

As always, we are deeply grateful to the authors of the submitted papers and to the reviewers (listed elsewhere in this volume) who produced three thorough and thoughtful reviews for each paper in a fairly short review period. The quality of submitted work continues to grow, and the organizers are truly grateful to the members of our amazing Program Committee, who helped us to determine which work was ready to be presented, and which would benefit from the additional experiments and analyses suggested by the reviewers. As in years past, we are looking forward to a productive workshop and

[1] https://www.coalitionforhealthai.org/

[2] https://www.fda.gov/media/153486/download

[3] https://tripod-llm.vercel.app/

hoping it will foster new collaborations and research. This will enable our community to continue making valuable contributions to public health and well-being and clinical research.

The advent of Generative AI and LLMs has also transformed our workshop introducing new challenges and opportunities. We are now in the era of BioMedGen.

Organizing Committee

Organizers:

Dina Demner-Fushman, US National Library of Medicine
Sophia Ananiadou, National Centre for Text Mining and University of Manchester, UK
Makoto Miwa, Toyota Technological Institute, Japan
Kirk Roberts, UTHealth, Houston, Texas, USA
Junichi Tsujii, National Institute of Advanced Industrial Science and Technology, Japan

Program Committee:

Joseph Damilola Akinyemi, University of Ibadan, Ibadan, Nigeria
Sophia Ananiadou, National Centre for Text Mining and University of Manchester, UK
Emilia Apostolova, Anthem, Inc.
Eiji Aramaki, University of Tokyo, Japan
Berry de Bruijn, National Research Council Canada
Leonardo Campillos Llanos, Universidad Autonoma de Madrid, Spain
Jiyu Chen, CSIRO Data61, Australia
Mike Conway, University of Utah, USA
Surabhi Datta, Melax Tech, USA Dina Demner-Fushman, US National Library of Medicine
Dmitriy Dligach, Loyola University Chicago, USA
Kathleen C. Fraser, National Research Council Canada
Yanjun Gao, University of Wisconsin-Madison, USA
Natalia Grabar, CNRS, France
Cyril Grouin, LIMSI - CNRS, France
Tudor Groza, The Garvan Institute of Medical Research, Australia
Deepak Gupta, US National Library of Medicine
Thierry Hamon, LIMSI-CNRS, France
Sam Henry, Christopher Newport University, USA
William Hogan, UCSD, USA
Brian Hur, University of Washington, USA
Richard Jackson, AstraZeneca
Antonio Jimeno Yepes, IBM, Melbourne Area, Australia
Sarvnaz Karimi, CSIRO, Australia
Nazmul Kazi, Montana State University, USA
Won Gyu Kim, US National Library of Medicine
Roman Klinger, University of Stuttgart, Germany
Anna Koroleva, Omdena Andre Lamurias, University of Lisbon, Portugal
Majid Latifi, University of York, York, UK
Alberto Lavelli, FBK-ICT, Italy
Robert Leaman, US National Library of Medicine
Lung-Hao Lee, National Central University, Taiwan
Ulf Leser, Humboldt Universitat zu Berlin, Germany
Diwakar Mahajan, MIT-IBM Watson AI Lab, USA
Timothy Miller, Boston Children's Hospital and Harvard Medical School, USA
Makoto Miwa, Toyota Technological Institute, Japan
Claire Nedellec, INRA, France
Guenter Neumann, DFKI, Saarland, Germany
Aurelie Neveol, LIMSI - CNRS, France
Mariana Neves, German Federal Institute for Risk Assessment, Germany
Brian Ondov, Yale University, USA

Laura Plaza, Universidad Nacional de Educacion a Distancia, Spain
Noon Pokaratsiri Goldstein, DFKI, Germany
Francois Remy, Ghent University, Belgium
Francisco J. Ribadas-Pena, University of Vigo, Spain
Roland Roller, DFKI GmbH, Berlin, Germany
Mourad Sarrouti, Sumitovant Biopharma, Inc., USA
Peng Su, University of Delaware, USA
Madhumita Sushil, University of California, San Francisco, USA
Mario Sanger, Humboldt Universitat zu Berlin, Germany
Andrew Taylor, Yale University School of Medicine, USA
Karin Verspoor, RMIT University, Australia
Davy Weissenbacher, Cedars-Sinai, Los Angeles, California, USA
Nathan M. White, James Cook University, Australia
W John Wilbur, US National Library of Medicine
Amelie Wuhrl, University of Stuttgart, Germany
Dongfang Xu, Cedars-Sinai, USA.
Shweta Yadav, University of Illinois Chicago, USA
Jingqing Zhang, Imperial College London, UK
Ayah Zirikly, Johns Hopkins, USA.
Pierre Zweigenbaum, LIMSI - CNRS, France

Table of Contents

Conference Program

Friday, August 16, 2024

08:15–08:30 Opening remarks

08:30–10:30 Session 1: Oral Presentations

08:30–08:50 *Improving Self-training with Prototypical Learning for Source-Free Domain Adaptation on Clinical Text*
Seiji Shimizu, Shuntaro Yada, Lisa Raithel and Eiji ARAMAKI

08:50–09:10 *Generation and Evaluation of Synthetic Endoscopy Free-Text Reports with Differential Privacy*
Agathe Zecevic, Xinyue Zhang, Sebastian Zeki and Angus Roberts

09:10–09:30 *Evaluating the Robustness of Adverse Drug Event Classification Models using Templates*
Dorothea MacPhail, David Harbecke, Lisa Raithel and Sebastian Möller

09:30–09:50 *Advancing Healthcare Automation: Multi-Agent System for Medical Necessity Justification*
Himanshu Gautam Pandey, Akhil Amod and Shivang Kumar

09:50–10:10 *Open (Clinical) LLMs are Sensitive to Instruction Phrasings*
Alberto Mario Ceballos-Arroyo, Monica Munnangi, Jiuding Sun, Karen Zhang, Jered McInerney, Byron C. Wallace and Silvio Amir

10:10–10:30 *Analysing zero-shot temporal relation extraction on clinical notes using temporal consistency*
Vasiliki Kougia, Anastasiia Sedova, Andreas Joseph Stephan, Klim Zaporojets and Benjamin Roth

10:30–11:00 ***Coffee Break***

Friday, August 16, 2024 (continued)

11:00–13:00 Session 2: Shared Tasks

11:00–11:20 *Overview of the First Shared Task on Clinical Text Generation: RRG24 and "Discharge Me!"*
Justin Xu, Zhihong Chen, Andrew Johnston, Louis Blankemeier, Maya Varma, Jason Hom, William J. Collins, Ankit Modi, Robert Lloyd, Benjamin Hopkins, Curtis Langlotz and Jean-Benoit Delbrouck

11:20–11:25 *e-Health CSIRO at RRG24: Entropy-Augmented Self-Critical Sequence Training for Radiology Report Generation*
Aaron Nicolson, Jinghui Liu, Jason Dowling, Anthony Nguyen and Bevan Koopman

11:25–11:30 *WisPerMed at "Discharge Me!": Advancing Text Generation in Healthcare with Large Language Models, Dynamic Expert Selection, and Priming Techniques on MIMIC-IV*
Hendrik Damm, Tabea Margareta Grace Pakull, Bahadır Eryılmaz, Helmut Becker, Ahmad Idrissi-Yaghir, Henning Schäfer, Sergej Schultenkämper and Christoph M. Friedrich

11:30–11:50 *Overview of the BioLaySumm 2024 Shared Task on the Lay Summarization of Biomedical Research Articles*
Tomas Goldsack, Carolina Scarton, Matthew Shardlow and Chenghua Lin

11:50–12:00 *UIUC_BioNLP at BioLaySumm: An Extract-then-Summarize Approach Augmented with Wikipedia Knowledge for Biomedical Lay Summarization*
Zhiwen You, Shruthan Radhakrishna, Shufan Ming and Halil Kilicoglu

12:00–12:20 ***Shared Tasks Discussion***

12:20–13:00 ***Invited Talk – Titipat Achakulvisut: Enhancing Neuroscience Conferences through Natural Language Processing***

13:00–14:30 ***Lunch***

Friday, August 16, 2024 (continued)

14:00–15:30 BioNLP Poster Session

End-to-End Relation Extraction of Pharmacokinetic Estimates from the Scientific Literature
Ferran Gonzalez Hernandez, Victoria Smith, Quang Nguyen, Palang Chotsiri, Thanaporn Wattanakul, José Antonio Cordero, Maria Rosa Ballester, Albert Sole, Gill Mundin, Watjana Lilaonitkul, Joseph F. Standing and Frank Kloprogge

KG-Rank: Enhancing Large Language Models for Medical QA with Knowledge Graphs and Ranking Techniques
Rui Yang, Haoran Liu, Edison Marrese-Taylor, Qingcheng Zeng, Yuhe Ke, Wanxin Li, Lechao Cheng, Qingyu Chen, James Caverlee, Yutaka Matsuo and Irene Li

MedExQA: Medical Question Answering Benchmark with Multiple Explanations
Yunsoo Kim, Jinge Wu, Yusuf Abdulle and Honghan Wu

Do Clinicians Know How to Prompt? The Need for Automatic Prompt Optimization Help in Clinical Note Generation
Zonghai Yao, Ahmed Jaafar, Beining Wang, Zhichao Yang and hong yu

Domain-specific or Uncertainty-aware models: Does it really make a difference for biomedical text classification?
Aman Sinha, Timothee Mickus, Marianne Clausel, Mathieu Constant and Xavier Coubez

Can Rule-Based Insights Enhance LLMs for Radiology Report Classification? Introducing the RadPrompt Methodology.
Panagiotis Fytas, Anna Breger, Ian Selby, Simon Baker, Shahab Shahipasand and Anna Korhonen

Using Large Language Models to Evaluate Biomedical Query-Focused Summarisation
Hashem Hijazi, Diego Molla, Vincent Nguyen and Sarvnaz Karimi

Continuous Predictive Modeling of Clinical Notes and ICD Codes in Patient Health Records
Mireia Hernandez Caralt, Clarence Boon Liang Ng and Marek Rei

Can GPT Redefine Medical Understanding? Evaluating GPT on Biomedical Machine Reading Comprehension
Shubham Vatsal and Ayush Singh

Get the Best out of 1B LLMs: Insights from Information Extraction on Clinical Documents
Saeed Farzi, Soumitra Ghosh, Alberto Lavelli and Bernardo Magnini

K-QA: A Real-World Medical Q&A Benchmark
Itay Manes, Naama Ronn, David Cohen, Ran Ilan Ber, Zehavi Horowitz-Kugler and Gabriel Stanovsky

Friday, August 16, 2024 (continued)

XrayGPT: Chest Radiographs Summarization using Large Medical Vision-Language Models
Omkar Chakradhar Thawakar, Abdelrahman M. Shaker, Sahal Shaji Mullappilly, Hisham Cholakkal, Rao Muhammad Anwer, Salman Khan, Jorma Laaksonen and Fahad Khan

Multilevel Analysis of Biomedical Domain Adaptation of Llama 2: What Matters the Most? A Case Study
Vicente Ivan Sanchez Carmona, Shanshan Jiang, Takeshi Suzuki and Bin Dong

Mention-Agnostic Information Extraction for Ontological Annotation of Biomedical Articles
Oumaima El Khettari, Noriki Nishida, Shanshan Liu, Rumana Ferdous Munne, Yuki Yamagata, Solen Quiniou, Samuel Chaffron and Yuji Matsumoto

Automatic Extraction of Disease Risk Factors from Medical Publications
Maxim Rubchinsky, Ella Rabinovich, Adi Shribman, Netanel Golan, Tali Sahar and Dorit Shweiki

Intervention extraction in preclinical animal studies of Alzheimer's Disease: Enhancing regex performance with language model-based filtering
YIYUAN PU, Kaitlyn Hair, Daniel Beck, Mike Conway, Malcolm MacLeod and Karin Verspoor

Efficient Biomedical Entity Linking: Clinical Text Standardization with Low-Resource Techniques
Akshit Achara, Sanand Sasidharan and Gagan N

XAI for Better Exploitation of Text in Medical Decision Support
Ajay Madhavan Ravichandran, Julianna Grune, Nils Feldhus, Aljoscha Burchardt, Roland Roller and Sebastian Möller

Optimizing Multimodal Large Language Models for Detection of Alcohol Advertisements via Adaptive Prompting
Daniel Cabrera Lozoya, Jiahe Liu, Simon D'Alfonso and Mike Conway

Extracting Epilepsy Patient Data with Llama 2
Ben Holgate, Shichao Fang, Anthony Shek, Matthew McWilliam, Pedro Viana, Joel S. Winston, James T. Teo and Mark P. Richardson

How do you know that? Teaching Generative Language Models to Reference Answers to Biomedical Questions
Bojana Bašaragin, Adela Ljajić, Darija Medvecki, Lorenzo Cassano, Miloš Košprdić and Nikola Milošević

Low Resource ICD Coding of Hospital Discharge Summaries
Ashton Williamson, David de Hilster, Amnon Meyers, Nina Hubig and Amy Apon

Towards ML-supported Triage Prediction in Real-World Emergency Room Scenarios
Faraz Maschhur, Klaus Netter, Sven Schmeier, Katrin Ostermann, Rimantas Palunis, Tobias Strapatsas and Roland Roller

Creating Ontology-annotated Corpora from Wikipedia for Medical Named-entity Recognition
Johann Frei and Frank Kramer

Paragraph Retrieval for Enhanced Question Answering in Clinical Documents
Vojtech Lanz and Pavel Pecina

15:30–16:00 ***Coffee Break***

16:00–17:50 **Shared Tasks Poster Session**

RRG24

CID at RRG24: Attempting in a Conditionally Initiated Decoding of Radiology Report Generation with Clinical Entities
Yuxiang Liao, Yuanbang Liang, Yipeng Qin, Hantao Liu and Irena Spasic

MAIRA at RRG24: A specialised large multimodal model for radiology report generation
Shaury Srivastav, Mercy Ranjit, Fernando Pérez-García, Kenza Bouzid, Shruthi Bannur, Daniel C. Castro, Anton Schwaighofer, Harshita Sharma, Maximilian Ilse, Valentina Salvatelli, Sam Bond-Taylor, Fabian Falck, Anja Thieme, Hannah Richardson, Matthew P. Lungren, Stephanie L. Hyland and Javier Alvarez-Valle

AIRI at RRG24: LLaVa with specialised encoder and decoder
Marina Munkhoeva, Dmitry Umerenkov and Valentin Samokhin

iHealth-Chile-1 at RRG24: In-context Learning and Finetuning of a Large Multimodal Model for Radiology Report Generation
Diego Campanini, Oscar Loch, Pablo Messina, Rafael Elberg and Denis Parra

iHealth-Chile-3&2 at RRG24: Template Based Report Generation
Oscar Loch, Pablo Messina, Rafael Elberg, Diego Campanini, Álvaro Soto, René Vidal and Denis Parra

Gla-AI4BioMed at RRG24: Visual Instruction-tuned Adaptation for Radiology Report Generation
Xi Zhang, Zaiqiao Meng, Jake Lever and Edmond S.L. Ho

SICAR at RRG2024: GPU Poor's Guide to Radiology Report Generation
Kiartnarin Udomlapsakul, Parinthapat Pengpun, Tossaporn Saengja, Kanyakorn Veerakanjana, Krittamate Tiankanon, Pitikorn Khlaisamniang, Pasit Supholkhan, Amrest Chinkamol, Pubordee Aussavavirojekul, Hirunkul Phimsiri, tara sripo, Chiraphat Boonnag, Trongtum Tongdee, Thanongchai Siriapisith, Pairash Saiviroonporn, Jiramet Kinchagawat and Piyalitt Ittichaiwong

Discharge Me!

Shimo Lab at "Discharge Me!": Discharge Summarization by Prompt-Driven Concatenation of Electronic Health Record Sections
Yunzhen He, Hiroaki Yamagiwa and Hidetoshi Shimodaira

Ixa-Med at Discharge Me! Retrieval-Assisted Generation for Streamlining Discharge Documentation
Jordan C. Koontz, Maite Oronoz and Alicia Pérez

QUB-Cirdan at "Discharge Me!": Zero shot discharge letter generation by open-source LLM
Rui Guo, Greg Farnan, Niall McLaughlin and Barry Devereux

e-Health CSIRO at "Discharge Me!" 2024: Generating Discharge Summary Sections with Fine-tuned Language Models
Jinghui Liu, Aaron Nicolson, Jason Dowling, Bevan Koopman and Anthony Nguyen

UF-HOBI at "Discharge Me!": A Hybrid Solution for Discharge Summary Generation Through Prompt-based Tuning of GatorTronGPT Models
Mengxian Lyu, Cheng Peng, Daniel Paredes, Ziyi Chen, Aokun Chen, Jiang Bian and Yonghui Wu

EPFL-MAKE at "Discharge Me!": An LLM System for Automatically Generating Discharge Summaries of Clinical Electronic Health Record
Haotian Wu, Paul Boulenger, Antonin Faure, Berta Céspedes, Farouk Boukil, Nastasia Morel, Zeming Chen and Antoine Bosselut

UoG Siephers at "Discharge Me!": Exploring Ways to Generate Synthetic Patient Notes From Multi-Part Electronic Health Records
Erlend Frayling, Jake Lever and Graham McDonald

Roux-lette at "Discharge Me!": Reducing EHR Chart Burden with a Simple, Scalable, Clinician-Driven AI Approach
Suzanne Wendelken, Anson Antony, Rajashekar Korutla, Bhanu Pachipala, James Shanahan and Walid Saba

Yale at "Discharge Me!": Evaluating Constrained Generation of Discharge Summaries with Unstructured and Structured Information
Vimig Socrates, Thomas Huang, Xuguang Ai, Soraya Fereydooni, Qingyu Chen, R Andrew Taylor and David Chartash

IgnitionInnovators at "Discharge Me!": Chain-of-Thought Instruction Finetuning Large Language Models for Discharge Summaries
An Quang Tang, Xiuzhen Zhang and Minh Ngoc Dinh

MLBMIKABR at "Discharge Me!": Concept Based Clinical Text Description Generation
Abir Naskar, Jane Hocking, Patty Chondros, Douglas Boyle and Mike Conway

BIoLaySum

DeakinNLP at BioLaySumm: Evaluating Fine-tuning Longformer and GPT-4 Prompting for Biomedical Lay Summarization
Huy Quoc To, Ming Liu and Guangyan Huang

ELiRF-VRAIN at BioLaySumm: Boosting Lay Summarization Systems Performance with Ranking Models
Vicent Ahuir, Diego Torres, Encarna Segarra and Lluís-F. Hurtado

BioLay_AK_SS at BioLaySumm: Domain Adaptation by Two-Stage Fine-Tuning of Large Language Models used for Biomedical Lay Summary Generation
Akanksha Karotia and Seba Susan

WisPerMed at BioLaySumm: Adapting Autoregressive Large Language Models for Lay Summarization of Scientific Articles
Tabea Margareta Grace Pakull, Hendrik Damm, Ahmad Idrissi-Yaghir, Henning Schäfer, Peter A. Horn and Christoph M. Friedrich

HULAT-UC3M at BiolaySumm: Adaptation of BioBART and Longformer models to summarizing biomedical documents
Adrian Gonzalez Sanchez and Paloma Martínez

Saama Technologies at BioLaySumm: Abstract based fine-tuned models with LoRA
Hwanmun Kim, Kamal raj Kanakarajan and Malaikannan Sankarasubbu

AUTH at BioLaySumm 2024: Bringing Scientific Content to Kids
Loukritia Stefanou, Tatiana Passali and Grigorios Tsoumakas

SINAI at BioLaySumm: Self-Play Fine-Tuning of Large Language Models for Biomedical Lay Summarisation
Mariia Chizhikova, Manuel Carlos Díaz-Galiano, L. Alfonso Ureña-López and María-Teresa Martín-Valdivia

RAG-RLRC-LaySum at BioLaySumm: Integrating Retrieval-Augmented Generation and Readability Control for Layman Summarization of Biomedical Texts
Yuelyu Ji, Zhuochun Li, Rui Meng, Sonish Sivarajkumar, Yanshan Wang, Zeshui Yu, Hui Ji, Yushui Han, Hanyu Zeng and Daqing He

Team YXZ at BioLaySumm: Adapting Large Language Models for Biomedical Lay Summarization
Jieli Zhou, Cheng Ye, Pengcheng Xu and Hongyi Xin

Eulerian at BioLaySumm: Preprocessing Over Abstract is All You Need
Satyam Modi and T Karthikeyan

HGP-NLP at BioLaySumm: Leveraging LoRA for Lay Summarization of Biomedical Research Articles using Seq2Seq Transformers
Hemang Malik, Gaurav Pradeep and Pratinav Seth

Ctyun AI at BioLaySumm: Enhancing Lay Summaries of Biomedical Articles Through Large Language Models and Data Augmentation
siyu bao, ruijing zhao, Siqin Zhang, jinghui zhang, weiyin wang and yunian ru

17:50–18:00 Closing remarks