BioNLP 2026: Program

Friday, July 3, 2026

08:40 - 08:50 **Opening remarks**

08:50 - 10:30 **Session 1: Clinical NLP**

08:50 - 09:00 **Trustworthy NLP for Low-Resource Languages: Agent-Based Uncertainty Modeling for Hebrew Radiology Report Structuring**
Hadas Ben Atya[1], Naama Gavrielov[2], Zvi Badash[3], Gili Focht[4], Ruth Cytter-Kuint[4], Talar Hagopian[5], Dan Turner[4], Moti Freiman[3]
[1]Technion, [2]Faculty of Biomedical Engineering, Technion, [3]Faculty of Biomedical Engineering, Technion -- Israel Institute of Technology, Haifa, Israel, [4]Juliet Keidan Institute of Pediatric Gastroenterology, Shaare Zedek Medical Center, Jerusalem, Israel, [5]Juliet Keidan Institute of Pediatric Gastroenterology, Shaare Zedek Medical Center, Jerusalem, Israel

09:00 - 09:10 **Reliable Automated Triage in Spanish Clinical Notes: A Hybrid Framework for Risk-Aware HIV Suspicion Identification**
Rodrigo Morales-Sánchez[1], Soto Montalvo[2], Raquel Martínez[3]
[1]Universidad Nacional de Educación a Distancia, [2]Universidad Rey Juan Carlos, [3]UNED

09:10 - 09:20 **Using Synthetic Records to Improve Automated Identification of Seizure Freedom in Clinical Text about People with Epilepsy**
Stephen Barlow[1], Yujian Gan[1], Joe Davies[1], Joel S Winston[2], James T Teo[1], Mark P Richardson[1], Ben Holgate[1]
[1]King's College London, [2]Kings College London

09:20 - 09:30 **A Multi-Agent Open-Source LLM for Structured Cancer Registry Information Extraction from Pathology and Medical Reports**
Abdulrahman AAl Abdulsalam[1], Adhari Al Zaabi[2], Riham Jeeballah[2], Habiba El keraby[2]
[1]University of Utah, [2]Sultan Qaboos University

09:30 - 09:40 **Clinical Evidence and Patient Reviews: A Linked Dataset for Antidepressant Side Effects**
Steven Au
UCSC

09:40 - 09:50 **Agentic AI Architectures for SOAP Note Generation**
Keno Hanken
Independent Researcher

09:50 - 10:00 **PromptRad: Knowledge-Enhanced Multi-Label Prompt-Tuning for Low-Resource Radiology Report Labeling**
Ying-Jia Lin[1], Tzu-Chin Lo[2], Ping-Chien Li[3], Chi-Tung Cheng[3], Chien-Hung Liao[3], Hung-Yu Kao[4]
[1]Chang Gung University, [2]Cathay General Hospital, [3]Chang Gung Memorial Hospital, [4]National Cheng Kung University

10:00 - 10:10 **When Demographic Sensitivity Isn't What It Seems: Baseline-Aware Counterfactual Audits for Clinical NLP**
Hyunwoo Yoo
Drexel University

10:10 - 10:20 **MAX-EVAL-11: A Large Scale Benchmark for Evaluating Large Language Models on Full-Spectrum ICD-11 Medical Coding**
Ujjwal Singh[1], Sarthak Deshwal[2], Nitish Dube[1], Arjun Sharma[1]
[1]Max Healthcare Institute Limited, [2]Plaid Inc.

10:20 - 10:30 **Reading Between the Lines: Toward Translating Verbose Patient-authored Messages into Clinician-Formulated Questions**
Sarvesh Soni, Madeline Bittner, Dina Demner-Fushman
National Library of Medicine

10:30 - 11:00 Coffee Break

11:00 - 11:30 Invited Talk -- Annika Marie Schoene, AI Safety in Healthcare : Ethical and Technical Considerations

11:30 - 12:30	**Session 2: Shared Tasks Overviews**
11:30 - 11:45	**Overview of the MedGenVidQA 2026 Shared Task on Medical Generative Video Question Answering** *Deepak Gupta[1], Collin S Campbell[2], Pedram Golnari[3], Dina Demner-Fushman[4]* [1]National Library of Medicine, NIH, [2]Rowan-Virtua School of Osteopathic Medicine, [3]Case Western Reserve University, [4]National Library of Medicine
11:45 - 12:00	**Overview of the Medical Decision Extraction, Analysis, and Classification Task (MedExACT) of BioNLP 2026** *Mohamed Elgaar[1], Jiali Cheng[2], Nidhi Vakil[3], Mehrnaz Sadrolashrafi[4], Mitra Mohtarami[5], Adrian Wong[4], Hadi Amiri[1], Leo Anthony Celi[6]* [1]University of Massachusetts Lowell, [2]Northeastern University, [3]University of Massachusetts, Lowell, [4]Beth Israel Deaconess Medical Center, [5]Anselm College, [6]Massachusetts Institute of Technology
12:00 - 12:15	**Overview of the PsyDefDetect Shared Task at BioNLP 2026: Detecting Levels of Psychological Defense Mechanisms in Supportive Conversations** *Hongbin Na[1], Zimu Wang[2], Zhaoming Chen[3], Yining Hua[4], Rena Gao[5], Kailai Yang[6], Ling Chen[1], Wei Wang[7], Shaoxiong Ji[8], John Torous[4], Sophia Ananiadou[6]* [1]University of Technology Sydney, [2]University of Liverpool, [3]University of Utah, [4]Harvard University, [5]The University of Melbourne, [6]University of Manchester, [7]Xi'an Jiaotong-Liverpool University, [8]University of Turku and ELLIS Institute Finland
12:15 - 12:30	**Overview of the ClinicalSkillQA 2026 Shared Task on Continuous Perception and Procedural Reasoning in Clinical Skill Assessment** *Xiyang Huang[1], Renxiong Wei[2], Yihuai Xu[3], Zhiyuan Chen[4], Keying Wu[3], Jiayi Xiang[3], Buzhou Tang[5], Yanqing Ye[2], Jinyu Chen[2], Cheng Zeng[3], Min Peng[4], Qianqian Xie[1], Sophia Ananiadou[6]* [1]Wuhan University, [2]Zhongnan Hospital of Wuhan University, [3]School of Artificial Intelligence, Wuhan University, [4]School of Computer Science, Wuhan University, [5]Harbin Institute of Technology (Shenzhen), [6]University of Manchester
12:30 - 14:00	Lunch
14:00 - 15:30	**Session 3: Foundational tasks**

14:00 - 14:10 **BioConflict: A Benchmark for Evaluating Large Language Models in Biomedical Contradiction Detection and Consensus Synthesis**
Ashwin Kirubakaran[1] and Henry Gagnier[2]
[1]Edison Academy Magnet School, [2]Pittsford Sutherland High School

14:10 - 14:20 **Hierarchy-Aware Hyperbolic and Semantic Reranking for Ontology-Based Phenotype Linking**
Thomas LABBE[1], Moussa BADDOUR[2], Axel BONESTEVE[3], Paul ROLLIER[3], Marie DE TAYRAC[3], Olivier DAMERON[4]
[1]Orange / Bcom / CHU Rennes, [2]bcom, [3]CHU Rennes, [4]Rennes University

14:20 - 14:30 **GRAFT: Gated Retrieval-Augmented Fine-Tuning for Relation Extraction**
Yuhang Jiang[1] and Ramakanth Kavuluru[2]
[1]University of Indiana, [2]University of Kentucky

14:30 - 14:40 **Can NLP Models Detect When One Publication Outweighs Twenty? Predicting Systematic Review Conclusion Changes**
Ebrahim Awad Alharbi[1] and Mark Stevenson[2]
[1]The University of Sheffield, [2]University of Sheffield

14:40 - 14:50 **Divide-Prompt-Refine: a Training-Free, Structure-Aware Framework for Biomedical Abstract Generation**
Sylvey Lin[1], Joseph D Menke[1], Shufan Ming[1], Dongin Nam[1], Neil Smalheiser[2], Halil Kilicoglu[3]
[1]University of Illinois Urbana-Champaign, [2]University of Illinois at Chicago, [3]University of Illinois at Urbana-Champaign

14:50 - 15:00 **What Do Biomedical NER and Entity Linking Benchmarks Measure? A Corpus-Centric Diagnostic Framework**
Robert Leaman[1], Rezarta Islamaj[2], Zhiyong Lu[3]
[1]National Center for Biotechnology Information, [2]National Center for Biotechnology Information, National Library of Medicine, [3]NCBI, NLM, NIH

15:30 - 16:00 Coffee Break

15:30 - 18:00 **Poster Session**

Agentic Feature Selection via LLM for Epileptic Seizure Detection
Aizierjiang Aiersilan and Xiaodong Qu
The George Washington University

Training Biomedical Retrievers From Large-Scale Citation Contexts
Xing David Wang[1], Duy Le Thanh[2], Ulf Leser[2]
[1]Computer Science Department, Humboldt-Universität zu Berlin, [2]Humboldt-Universität zu Berlin

Gold Label Errors in the SciFact Benchmark: An LLM-Assisted Annotation Audit
Julien Sylvestre
BXCT Partners

BioRAG: A Systematic Ablation Study of Retrieval Strategies for Biomedical Question Answering
Krushil Bhojani, Mayank Waghmare, Hima Bindu Nandyala
Suny Polytechnic Institute

Post Hoc Agentic Refinement for Improving Precision in Multilingual Clinical Text De-identification
Justin Xu[1], Alistair EW Johnson[2], Thomas Lin[2], David W Eyre[1], Rodolfo Quispe[2]
[1]University of Oxford, [2]Microsoft Health & Life Sciences

Do Syntactic Features Help Biomedical Relation Extraction? An Empirical Study of Verb Token and Dependency Graph Augmentation
Mustafa K Sikder and Ernest Kwegyir-Afful
U.S. Food and Drug Administration

Beyond Knowledge Graphs: PubMedBERT Embeddings as a Competitive Standalone Modality for Drug Re-purposing
Rishik K Kondadadi[1] and John E. Ortega[2]
[1]Student, [2]Northeastern University

CoreELM: An Open-Source Framework for Aligning Large Language Models to Embedding Spaces
Brian Ondov[1], Chia-Hsuan Chang[2], Yujia Zhou[1], Mauro Giuffrè[1], Hua Xu[1]
[1]Yale School of Medicine, [2]Yale University

Uncertainty-Aware Multi-Label Routing of Clinical Text to Surveillance Pathways

Agathe Zecevic[1], Sebastian S Zeki[2], Angus Roberts[3]
[1]King' College London, [2]Guy's and St Thomas' NHS Trust, [3]King's College London

MedCAT v2: a modular, extensible architecture for clinical named entity recognition and linking under real-world privacy and compute constraints

Mart Ratas, Thomas Searle, Adam Sutton, Richard Dobson
King's College London

Effects of Adaptive Pretraining in Specialized Domains for Named Entity Recognition

Jack Lynam[1] and Sam Henry[2]
[1]Christopher Newport University, [2]Randolph Macon College

Trade-offs in Medical LLM Adaptation: An Empirical Study in French QA

Ikram Belmadani[1], Oumaima El Khettari[2], Carlos Ramisch[3], Frederic Bechet[4], Richard Dufour[5], Benoit Favre[6]
[1]Aix-Marseille University, [2]Nantes Université - LS2N, [3]Aix Marseille University, CNRS, LIS, [4]Aix Marseille Universite - LIS/CNRS, [5]LS2N - Nantes University, [6]Aix-Marseille University LIS/CNRS

Diagnosing Lower Extremity Arteriovenous Diseases Using Agentic LLMs

Zicen Liao[1], Yunhao Sun[2], Matthew Purver[3]
[1]+44 07881729773, [2]sunyh25@alumni.sysu.edu.cn, [3]m.purver@qmul.ac.uk

KGRxn-LLM: Knowledge Graph Enhanced Large Language Models for Molecular Reaction Reasoning

Weichen Liu[1], Qiyao Xue[1], Yuyang Wu[2], Olexandr Isayev[2], Natasa Miskov-Zivanov[3]
[1]University of Pittsburgh, [2]Carnegie Mellon University, [3]ECE, BioE, CompBio, Univ. of Pittsburgh

Treating Decoder-Only LLMs as Encoders: A Simple and Effective Fine-tuning Approach for Named Entity Recognition

Ken Yano[1] and Hiroya Takamura[2]
[1]The National Institute of Advanced Industrial Science and Technology, [2]The National Institute of Advanced Industrial Science and Technology (AIST)

A Multi-View Framework for Cross-Domain Nutrition Misinformation Detection in Social Media

Vishwaa Shah, Indika Kahanda, Andrea Arikawa, Asal Abbaszadeh, Richard Loftis
University of North Florida

Ontological Validation of Biomedical Topic Models: SNOMED CT Hierarchy Distance as an Automated Evaluation Metric
Ilan Rubinfeld, Sami Zaidi, Milosh Djuric, Loay Kabbani, Mouhammad Halabi, Alex Shepard
Henry Ford Health

Systematic Evaluation of the Quality of Synthetic Clinical Notes Rephrased by LLMs at Million-Note Scale
Jinghui Liu[1], Sarvesh Soni[2], Anthony Nguyen[1]
[1]CSIRO, [2]NIH

Bridging the Version Gap: Multi-version Training Improves ICD Code Prediction, Especially for Rare Codes
Jinghui Liu and Anthony Nguyen
CSIRO

EmCellLLM: Human Peri-Implantation Embryonic Cell Annotation Based on Large Language Models
Xiaorui Guo[1], Zhiwei Liu[2], Qianqian Xie[3], Sophia Ananiadou[2]
[1]The University of Edinburgh, [2]University of Manchester, [3]Wuhan University

Citation-Aware Continual Pre-Training for Biomedical Language Models
Masaki Asada[1], Tomoki Tsujimura[1], Tatsuya Ishigaki[2], Shusaku Egami[2], Ken Fukuda[3], Hiroya Takamura[4]
[1]National Institute of Advanced Industrial Science and Technology, [2]National Institute of Advanced Industrial Science and Technology (AIST), [3]AIRC/AIST, [4]The National Institute of Advanced Industrial Science and Technology (AIST)

TrackList: Tracing Back Query Linguistic Diversity for Head and Tail Medical Knowledge in Open Large Language Models
Ioana Buhnila[1], Aman Sinha[2], Mathieu Constant[3]
[1]Center for Data Science in Humanities, Chosun University, [2]University of Lorraine, [3]Université de Lorraine, CNRS, ATILF

Discharge Instructions are not One Task: Grounding Differences Between Surgical and Non-Surgical Admissions
Mayank Jobanputra[1], Justin Xu[2], Samarth H Oza[3], Hulma Naseer[1], Yifan Wang[1], Blerta Veseli[1], Chandralekha Kona[4], Haochen Cui[2], David W Eyre[2], Vera Demberg[1]

[1]Saarland University, [2]University of Oxford, [3]All India Institute of Medical Sciences Rajkot, [4]Independent resaercher

PrionNER: A Named Entity Recognition Dataset for Prion Disease Biomedical Literature
An Tuan Dao[1], Nhan Hoang Tri Ly[2], Thao Thi Thu Tran[2], Yuji Matsumoto[3], Akiko Aizawa[4]
[1]The University of Tokyo, [2]Medical Doctor, Independent Researcher, [3]Tohoku University, [4]National Institute of Informatics

BioTopicXplor: A Web Tool for Interactive Exploration of PubMed Literature through Reproducible Topics.
Lana Yeganova[1], Donald C Comeau[2], Won Kim[2], Natalie Xie[2], Shubo Tian[2], W John Wilbur[3], Zhiyong Lu[4]
[1]NIH, [2]NLM/NIH, [3]NCBI, [4]NCBI, NLM, NIH

Investigating Stigmatizing Language in Clinical Documentation with Open-Source Large Language Models
Rajashree Dahal, Pardis Hosseinpour, Pranithi Kamishetty, Satwik Pamulaparthy, Saeid Tizpaz-Niari, Natalie Parde
University of Illinois Chicago

Learning to Combine AI Annotations for Improved Biomedical Relevance Labeling
Won Gyu KIM[1], Lana Yeganova[2], Shubo Tian[3], Donald C Comeau[3], W John Wilbur[4], Zhiyong Lu[5]
[1]DIR/NLM/NIH, [2]NIH, [3]NLM/NIH, [4]NCBI, [5]NCBI, NLM, NIH

When Does Retrieval Beat Direct LLM Diagnosis in Rare Disease? An Empirical Study of Ontology Coverage
Mohamed Elmofty[1] and Ulf Leser[2]
[1]Humboldt University of Berlin, [2]Humboldt-Universität zu Berlin

BioCoref: Benchmarking Biomedical Coreference Resolution with LLMs
Nourah M Salem[1], Elizabeth White[2], Michael Bada[2], Lawrence Hunter[2]
[1]Computational Bioscience Program, University of Colorado, Anschutz Medical Campus, Aurora, CO, 80045, USA, [2]University of Chicago

Tokenization Granularity and Medical Term Representations in Language Models
Vojtech Lanz and Pavel Pecina
Charles University

CAP: A Source-Grounded Proposition Scaffold for Faithful Clinical Dialogue-to-Note Generation

Hyunkyung Lee, Jisoo Jung, Jeonguk Lee, Jaehyo Yoo, Wooseok Han, Minkyu Kim, Gibaeg Kim
AITRICS

Segmentation Matters: Exploring LLM-Based Strategies for Temporal Clinical Event Identification in Oncology Reports

Cristiano Bellucci[1], Francesco Madeddu[2], Chiara Iacomini[3], Carlotta Masciocchi[3], Stefano Patarnello[3], Massimo Bernaschi[2], Mario Santoro[2], Livia Lilli[4]
[1]KeyBiz srl, [2]Istituto per le Applicazioni del Calcolo "Mauro Picone", Italian National Research Council, [3]Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy, [4]Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy; Catholic University of the Sacred Heart, Rome, Italy

Operation-Mechanism Alignment for Reliable Clinical Reasoning over Electronic Health Records

Guanyu Tao[1], Siyao Wang[2], Yong Xue[1], Ashwani Tanwar[1], Yuting Ji[1], Kai Sun[2], Monica Mok[1], Marzana Chowdhury[1], Deepa Gupta[1], Ashok Gupta[1], Jingqing Zhang[1], Vibhor Gupta[1], Yike Guo[3]
[1]Pangaea Data Limited, UK, USA, [2]Pangaea Data Limited, UK, USA; Data Science Institute, Imperial College London, London, UK, [3]Pangaea Data Limited, UK, USA; Hong Kong University of Science and Technology, Hong Kong SAR, China

A Comparative Analysis of In-Context Learning and Fine-Tuning for Biomedical Information Retrieval and Sentence Extraction Using Research Domain Criteria

Athlene Jones, Khanh Linh Lieu, Indika Kahanda
University of North Florida

A Deterministic Multi-Stage Retrieval Pipeline for Longitudinal EHR Question Answering

Shubham Agarwal[1], Thomas Searle[1], Richard Dobson[1], Ninoslav Majkic[2], Niko Moller-Grell[1]
[1]King's College London, [2]South London and Maudsley NHS Foundation Trust

Interpretable ICD Code Classification with Faithful Sentence Extraction

Yichen Wang[1], Lian Hong[1], Masato Mizogaki[2], Shunnosuke Umeda[1], Toshinune Kenmotsu[1], Akihiro Tamura[3], Daniel Andrade[1]
[1]Hiroshima University, [2]Bourbon Corporation, [3]Doshisha University

Evaluating LLM-as-a-Judge for Medical Term Simplification
Ioana Buhnila[1], Aman Sinha[2], Rohit Agarwal[3], Dilip K. Prasad[3], Mathieu Constant[4]
[1]Center for Data Science in Humanities, Chosun University, [2]University of Lorraine, [3]UiT The Arctic University of Norway, [4]Université de Lorraine, CNRS, ATILF

FACT: Functional Group Alignment and Consistency in Token Space for Structure-aware Molecular Representation Learning
Hyeonyeong Nam[1], Woojae Choi[1], Deok-Joong Lee[2], Young-Han Son[1], Sangwoon Lee[3], Bogyeong Kang[1], Eunjung Jo[1], Tae-Eui Kam[1]
[1]Korea University, [2]Department of Artificial Intelligence, Korea University, [3]Korea University, Department of Artificial Intelligence

Diagnosable ColBERT: Debugging Late-Interaction Retrieval Models Using a Learned Latent Space as Reference
François Remy
Parallia AI

Developing Literature Annotation Guidelines for Representing Normal Physiology in Biolink-Compatible Knowledge Graphs
Madeline Bittner, Willie Rogers, Dina Demner-Fushman, Richard H Scheuermann, Matthew Diller
National Library of Medicine

CENT: Context Engineering Framework for Normalization of Clinical Trial Procedures
Sanya Bathla Taneja[1], Ziqing Ji[2], Hans Verstraete[1], Ali Samadani[1]
[1]Johnson and Johnson Innovative Medicine, [2]University of Washington

VERICITE: Evaluating Sentence-Level Citation Faithfulness in Retrieval-Augmented Medical Question Answering
Yixian Ma[1], Bohao Chu[2], Norbert Fuhr[3]
[1]University of Duisburg Essen, [2]Universisty Duisburg-Essen, [3]University of Duisburg-Essen

AAbAAC: An Annotated Corpus for Autoimmunity Information Extraction
Fabien Maury[1], Solène Grosdidier[2], Maud de Dieuleveult[1], Adrien Coulet[3]
[1]Inserm, [2]Freelance researcher, [3]Inria

Towards Grounded Hallucination Definitions for Biomedical Question Answering with Reproducible Examples from REMOVED FOR REVIEW
brandon C colelough[1], Davis Bartels[2], Madeline Bittner[3], Dina Demner-Fushman[3]
[1]NIH, [2]National Institutes of Health, [3]National Library of Medicine

VaxScope: Document-Level Structured Evidence Extraction from Immunization Systematic Reviews
Bahar Ilgen, Ebenezer Awotoro, Georges Hattab
Robert Koch Institute

Medical Context Variation: A source of impairment for Event classification
Aman Sinha[1], Marianne Clausel[1], Mathieu Constant[2], Xavier Coubez[3]
[1]University of Lorraine, [2]Université de Lorraine, CNRS, ATILF, [3]ICANS

KALIMBA: Knowledge-Assisted Literature Mining for Biological Interaction Analysis
Niloofar Arazkhani[1], Maciej K Kotecki[2], Brent Cochran[2], Natasa Miskov-Zivanov[3]
[1]University of Pittsburgh, [2]Department of Developmental, Molecular and Chemical Biology, Tufts University School of Medicine, Boston, [3]ECE, BioE, CompBio, Univ. of Pittsburgh

When Retrieval Doesn't Help: A Large-Scale Study of Biomedical RAG
Erfan Nourbakhsh, Rocky Slavin, Ke Yang, Anthony Rios
University of Texas at San Antonio

CrossDDI: Cross-Source Evidence-Grounded Drug-Drug Interaction Verification
Bohao Chu[1] and Norbert Fuhr[2]
[1]Universisty Duisburg-Essen, [2]University of Duisburg-Essen

SCoPE: Planning for Hybrid Querying over Clinical Trial Data
Suparno Roy Chowdhury[1], Manan Roy Choudhury[1], Tejas Anvekar[1], Muhammed Ali Khan[2], Kaneez Zahra Rubab Khakwani[2], Mohamad Bassam Sonbol[2], Irbaz Bin Riaz[2], Vivek Gupta[1]
[1]Arizona State University, [2]Mayo Clinic

Expert-Guided Schema-Based Structured Extraction from CONSORT Diagrams Using Vision-Language Models
Damian Stachura, Bartosz Przechera, Monika Opałek, Ewelina Sadowska, Ewa Borowiack, Artur Nowak
Evidence Prime

From Rules to Predictions: Federated Tabular Learning with LLM Reasoning
Afsaneh Mahanipour and Hana Khamfroush
Department of Computer Science, University of Kentucky

MedBench: Deliberative Evaluation of Medical Language Models
Pratik Jalan[1], Mukul Joshi[2], Akhilesh Magotra[3], Kshitij Jadhav[4]
[1]Indian Institute of Technology, Mumbai, [2]Mr., [3]Birla Institute of Technology & Science Pilani, [4]Indian Institute of Technology Bombay

Fast, Accurate, and Local Conversion of MIMIC-IV to OMOP with DBT
Adam Sutton, Niko Möller-Grell, Thomas Searle, Richard J Dobson
King's College London

Exploring Novel Drug Research Area using Large Language Models Based on Research Trends in Biomedical Literature
Afnan Afnan, Michael Van B Supranes, Tomohiro Nishiyama, Shoko Wakamiya, Eiji Aramaki
Nara Institute of Science and Technology

FHexchange: Resources for Family Health History Extraction and Normalization From Consumer Dialog Sources
Michelle H. Nguyen[1], Nidhi Soley[1], Ayah Zirikly[1], João Sedoc[2], Casey O. Taylor[1]
[1]Johns Hopkins University, [2]New York University

IndicMedDialog: A Parallel Multi-Turn Medical Dialogue Dataset for Accessible Healthcare in Indic Languages
Shubham Kumar Nigam[1], Suparnojit Sarkar[2], Piyush Patel[3]
[1]University of Birmingham, [2]Heritage Institute of Technology Kolkata, [3]Madan Mohan Malaviya University of Technology

Towards a Radiologist Imitation Framework for 3D CT Diagnosis with Multimodal LLMs
Kaidi Zhang[1], Zhiyuan Yan[2], Gao Cheng[1], Zhenyang Cai[1]
[1]The Chinese University of Hong Kong, Shenzhen, [2]Shandong Second Medical University

Probing and Steering Uncertainty in Biomedical Language Models: Representational Structure and Behavioral Limits
Debmalya Pal
University of California, San Diego

Relations of Linguistic Features and Medical Text Preferences are Nontrivial
Davis Bartels[1], brandon C colelough[2], Dina Demner-Fushman[3]
[1]National Institutes of Health, [2]NIH, [3]National Library of Medicine

The Divergence Hypothesis: Unmasking Lexical Interference and Label Bias in Mental Health NLP
MOUSTAFA Yehia HASSAN
Doha Institute for Graduate Studies

Towards Unified Factuality Evaluation for Biomedical QA and Summarization: Aligning Metrics with Clinical Use-Cases
Mahule Roy[1] and Subhas Roy[2]
[1]University of Oxford, [2]TATA Consumer Products Limited

Analyzing Prompt Design Choices in Biomedical Information Extraction for Low-Resource Languages
Ayesha Khatun, Kadir Bulut Ozler, Steven Bethard, Egoitz Laparra
University of Arizona

Randomized Controlled Trials as the Gold-Standard for Evaluating LLMs: A Primer for Biomedical NLP Researchers
Vicente Ivan Sanchez Carmona[1], Shanshan Jiang[2], Bin Dong[1]
[1]Ricoh Software Research Center Beijing, [2]Ricoh Software Research Center (Beijing) Co., Ltd.

Evaluation of Multilingual Text Simplification for the Mental Health Domain: Exploring Small Language Models
Olga Pelloni[1], Sandra Anna Just[2], Lars Ailo Bongo[1]
[1]UiT The Arctic University of Norway, [2]UiT - The Arctic University of Norway

MeSHClass-ES and AnatEM-ES: Open Resources for Spanish Biomedical NLP
Santiago Martinez Novoa, Lina Maria Gomez Mesa, Juan Camilo Prieto, Ruben Francisco Manrique
Universidad de los Andes

Small LLMs for Biomedical Claim Verification: Cost-Effective Fine-Tuning, Structural Dataset Shortcuts, and Cross-Domain Generalization

Gaurav Kumar
University of California San Diego

Forgotten Words: Benchmarking NeoBERT for Dementia Detection in Low-Resource Conversational Filipino and English Speech

Rez Samantha Zamora Floresca[1], Edric Castel Chua Hao[2], Hannah Grachiella Buñales[1], Chelsea Dominique E. Temprosa[1], Georgianna Z. Reyes[1], Kervin Gabriel L. Chua[1]
[1]Ateneo De Manila Senior High School, [2]Independent