

# Dmitriy Dligach, Ph.D.

Loyola University Chicago  
Department of Computer Science  
1052 W. Loyola Avenue  
Chicago, IL 60626

Phone: (773) 508-3557  
Fax: (773) 508-7099  
Email: [dd@cs.luc.edu](mailto:dd@cs.luc.edu)  
Web: <http://www.dmitriydligach.com>

## Research Interests

Natural language processing, machine learning, artificial intelligence, data mining, medical informatics.

## Education

Ph.D. Computer Science, University of Colorado, Boulder, 2010.

M.S. Computer Science, State University of New York, Buffalo, 2004.

B.S. Computer Science, Loyola University, Chicago, 1999.

## Professional Experience

Associate Professor, Loyola University Chicago, 2021 - present.

Associate Professor, Loyola University Stritch School of Medicine, 2021 - present.

Honorary Fellow, University of Wisconsin School of Medicine, 2020 - present.

Assistant Professor, Loyola University Chicago, 2016 - 2021.

Assistant Professor, Loyola University Stritch School of Medicine, 2016 - 2021.

Instructor (entry-level faculty appointment), Harvard Medical School, 2014 - 2015.

Associate Scientific Research Staff, Boston Children's Hospital, 2014 - 2015.

Research Fellow, Harvard Medical School, 2011 - 2014.

Research Fellow, Boston Children's Hospital, 2011 - 2014.

Postdoctoral Research Associate, University of Colorado, Boulder, 2010 - 2011.

Research Assistant, University of Colorado, Boulder, 2005 - 2010.

Research Scientist, Pronto Inc., Boulder, CO, 2005.

Teaching Assistant, State University of New York, Buffalo, 2003 - 2005.

Senior Software Engineer, Motorola, Inc., Arlington Heights, IL, 2000 - 2003.

Development Analyst, Commonwealth Edison, Chicago, IL, 1999 - 2000.

Network Administrator, Graver, Bokhof, Goodwin & Sullivan, Chicago, IL, 1997 - 1998.

## Peer-reviewed Publications

### Journal Articles

Gallifant, Jack and Afshar, Majid and Ameen, Saleem and Aphinyanaphongs, Yindalon and Chen, Shan and Cacciamani, Giovanni and Demner-Fushman, Dina and Dligach, Dmitriy and Daneshjou, Roxana and Fernandes, Chrystinne and Hansen, Lasse Hyldig and Landman, Adam and Lehmann, Lisa and McCoy, Liam G. and Miller, Timothy and Moreno, Amy and Munch, Nikolaj and Restrepo, David and Savova, Guergana and Umeton, Renato and Gichoya, Judy Wawira and Collins, Gary S. and Moons, Karel G. M. and Celi, Leo A. and Bitterman, Danielle S. The TRIPOD-LLM Statement: A Targeted Guideline For Reporting Large Language Models Use. *Nature Medicine*. 2024.

Skatje Myers, Timothy A. Miller, Yanjun Gao, Matthew M. Churpek, Anoop Mayampurath, Dmitriy Dligach, Majid Afshar. Lessons Learned on Information Retrieval in Electronic Health Records: A Comparison of Embedding Models and Pooling Strategies. *Journal of American Medical Informatics Association*. 2024.

Yanjun Gao, Ruizhe Li, Emma Croxford, John R Caskey, Brian W Patterson, Matthew M. Churpek, Timothy Miller, Dmitriy Dligach, Majid Afshar. Leveraging A Medical Knowledge Graph into Large Language Models for Diagnosis Prediction: Design and Application Study. *Journal of Medical Internet Research (JMIR AI)*. 2024.

WonJin Yoon, Shan Chen, Yanjun Gao, Zhazhan Zhao, Dmitriy Dligach, Danielle S. Bitterman, Majid Afshar, Timothy Miller. LCD Benchmark: Long Clinical Document Benchmark on Mortality Prediction for Language Models. *Journal of American Medical Informatics Association*. 2024.

Yanjun Gao, Skatje Myers, Shan Chen, Dmitriy Dligach, Timothy A Miller, Danielle Bitterman, Matthew Churpek, Majid Afshar. When Raw Data Prevails: Are Large Language Model Embeddings Effective in Numerical Data Representation for Medical Machine Learning Applications? *Findings of Empirical Methods in Natural Language Processing*. 2024.

Pui Man Rosalind Lai, Elliot Akama-Garren, Anil Can, Selena-Rae Tirado, Victor M Castro, Dmitriy Dligach, Sean Finan, Vivian S Gainer, Nancy A Shadick, Guergana Savova, Shawn N Murphy, Tianxi Cai, Scott T Weiss, Rose Du. Family history as the strongest predictor of aortic and peripheral aneurysms in patients with intracranial aneurysms. *Journal of Clinical Neuroscience*. 2024.

Jifan Gao, Guanhua Chen, Ann P. O'Rourke, John Caskey, Kyle Carey, Madeline Oguss, Anne Stey, Dmitriy Dligach, Timothy Miller, Anoop Mayampurath, Matthew M. Churpek, Majid Afshar. Automated stratification of trauma injury severity across multiple body regions using multi-modal, multi-class machine learning models. *Journal of American Medical Informatics Association*. 2024.

Karway GK, Koyner JL, Caskey J, Spicer AB, Carey KA, Gilbert ER, Dligach D, Mayampurath A, Afshar M, Churpek M. Development and External Validation of Multimodal Postoperative Acute Kidney Injury Risk Machine Learning Models. *Journal of American Medical Informatics Association*. 2023.

Majid Afshar, Sabrina Adelaine, Felice Resnik, Marlon P Mundt, John Long, Margaret Leaf, Theodore Ampian, Graham J Wills, Benjamin Schnapp, Michael Chao, Randy Brown, Cara Joyce, Brihat Sharma, Dmitriy Dligach, Elizabeth S Burnside, Jane Mahoney, Matthew M Churpek, Brian W Patterson, Frank Liao. Deployment of Real-time Natural Language Processing and Deep Learning Clinical Decision Support in the Electronic Health Record: Pipeline Implementation and Protocol for an Opioid Misuse Screener in Hospitalized Adults. *JMIR Medical Informatics*. 2023.

Yanjun Gao, Dmitriy Dligach, Timothy Miller, Matthew M Churpek, Ozlem Uzuner, Majid Afshar. Progress Note Understanding—Assessment and Plan Reasoning: Overview of the 2022 N2C2 Track 3 shared task. *Journal of Biomedical Informatics*. 2023.

YanJun Gao, Dmitriy Dligach, Timothy Miller, John Caskey, Brihat Sharma, Matthew M Churpek, Majid Afshar. DR.BENCH: Diagnostic Reasoning Benchmark for Clinical Natural Language Processing. *Journal of Biomedical Informatics*. 2023. **Selected by AMIA 2023 as one of the top papers in 2023; featured in AMIA Year In Review.**

Cara Joyce, Talar W Markossian, Jenna Nikolaidis, Elisabeth Ramsey, Hale M Thompson, Juan C Rojas, Brihat Sharma, Dmitriy Dligach, Madeline K Oguss, Richard S Cooper, Majid Afshar. The Evaluation of a Clinical Decision Support Tool Using Natural Language Processing to Screen Hospitalized Adults for Unhealthy Substance Use: Protocol for a Quasi-Experimental Design. *JMIR Research Protocols*. 2022.

YanJun Gao, Dmitriy Dligach, Leslie Christensen, Samuel Tesch, Ryan Laffin, Dongfang Xu, Timothy Miller, Ozlem Uzuner, Matthew M Churpek, Majid Afshar. A Scoping Review of Publicly Available Language Tasks in Clinical Natural Language Processing. *Journal of American Medical Informatics Association*. 2022.

Majid Afshar, Brihat Sharma, Dmitriy Dligach, Madeline Oguss, Randall Brown, Neeraj Chhabra, Hale M Thompson, Talar Markossian, Cara Joyce, Matthew M Churpek, Niranjana S Karnik. Development and multimodal validation of a substance misuse algorithm for referral to treatment using artificial intelligence (SMART-AI): a retrospective deep learning study. *The Lancet Digit Health*. 2022.

John Caskey, Iain L. McConnell, Madeline Oguss, Dmitriy Dligach, Rachel Kulikoff, Brittany Grogan, Crystal Gibson, Elizabeth Wimmer, Traci E DeSalvo, Edwin E Nyakoe-Nyasani, Matthew M Churpek, Majid Afshar. A Natural Language Processing Pipeline to Identify COVID-19 Outbreaks from Contact Tracing Interview Forms for Public Health Departments. *JMIR Public Health and Surveillance*. 2022.

Jian Zhang, Anil Can, Pui Man Rosalind Lai, Srinivasan Mukundan, Victor M. Castro, Dmitriy Dligach, Sean Finan, Vivian S. Gainer, Nancy A. Shadick, Guergana Savova, Shawn N. Murphy, Tianxi Cai, Scott T. Weiss, Rose. Du. Geometric Features Associated with Middle Cerebral Artery Bifurcation Aneurysm Formation: A Matched Case-Control Study. *Journal of Stroke and Cerebrovascular Diseases*. 2022.

Hale M Thompson, Brihat Sharma, Sameer Bhalla, Randy Boley, Connor McCluskey, Dmitriy Dligach, Matthew M Churpek, Niranjana S Karnik, Majid Afshar. Bias and fairness assessment of a natural language processing opioid misuse classifier: detection and mitigation of electronic health record data disadvantages across racial subgroups. *Journal of American Medical Informatics Association*. 2021.

Kulshrestha S, Dligach D, Joyce C, Baker MS, Gonzalez R, O'Rourke AP, Glazer JM, Stey A, Kruser JM, Churpek MM, Afshar M. Comparison and Interpretability of Machine Learning Models to Predict Severity of Injury. *JAMIA Open*. 2021.

Jian Zhang, Pui Man Rosalind Lai, Anil Can, Srinivasan Mukundan, Victor M. Castro, Dmitriy Dligach, Sean Finan, Vivian S. Gainer, Nancy A. Shadick, Guergana Savova, Shawn N. Murphy, Tianxi Cai, Scott T. Weiss, and Rose Du. Tobacco use and age are associated with different morphologic features of anterior communicating artery aneurysms. *Scientific Reports*. 2021.

Jian Zhang, Anil Can, Pui Man Rosalind Lai, Srinivasan Mukundan Jr, Victor M Castro, Dmitriy Dligach, Sean Finan, Vivian S Gainer, Nancy A Shadick, Guergana Savova, Shawn N Murphy, Tianxi Cai, Scott T Weiss, Rose Du. Morphological Variables Associated with Ruptured Basilar Tip Aneurysms. *Scientific Reports*. 2021.

Dmitriy Dligach, Majid Afshar, Timothy Miller. Pre-training phenotyping classifiers. *Journal of Biomedical Informatics*. 2020.

Jian Zhang, Anil Can, Pui Man Rosalind Lai, Srinivasan Mukundan, Victor M Castro, Dmitriy Dligach, Sean Finan, Vivian Gainer, Nancy Shadick, Guergana Savova, Shawn N Murphy, Tianxi Cai, Scott T Weiss, Rose Du. Geometric Variations Associated with Posterior Communicating Artery Aneurysms. *Journal of NeuroInterventional Surgery*. 2020.

Jian Zhang, Anil Can, Pui Man Rosalind Lai, Srinivasan Mukundan Jr, Victor M Castro, Dmitriy Dligach, Sean Finan, Vivian S Gainer, Nancy A Shadick, Guergana Savova, Shawn N Murphy, Tianxi Cai, Scott T Weiss, Rose Du. Vascular geometry associated with anterior communicating artery aneurysm formation. *World Neurosurgery*. 2020.

Sujay Kulshrestha, Dmitriy Dligach, Cara Joyce, Marshall S Baker, Richard Gonzalez, Ann P O'Rourke, Joshua M Glazer, Anne Stey, Jacqueline M Kruser, Matthew M Churpek, Majid Afshar. Prediction of severe chest injury using natural language processing from the electronic health record. *Injury*. 2020.

Jian Zhang, Anil Can, Pui Man Rosalind Lai, Srinivasan Mukundan Jr, Victor M. Castro, Dmitriy Dligach, Sean Finan, Vivian S. Gainer, Nancy A. Shadick, Guergana Savova, Shawn N. Murphy, Tianxi Cai, Scott T. Weiss, and Rose Du. Surrounding Vascular Geometry Associated with Basilar Tip Aneurysm Formation. *Scientific Reports*. 2020.

Anoop Mayampurath, Matthew Churpek, Xin Su, Sameep Shah, Elizabeth Munroe, Bhakti Patel, Dmitriy Dligach, and Majid Afshar. External Validation of an Acute Respiratory Distress Syndrome Prediction Model Using Radiology Reports. *Critical Care Medicine*. 2020.

Brihat Sharma, Dmitriy Dligach, Kristin Swope, Elizabeth Salisbury-Afshar, Niranjana S. Karnik, Cara Joyce, Majid Afshar. Publicly available machine learning models for identifying opioid misuse from the clinical notes of hospitalized patients. *BMC Medical Informatics and Decision Making*. 2020.

Chen Lin, Steven Bethard, Dmitriy Dligach, Farig Sadeque, Guergana Savova, Timothy Miller. Does BERT Need Domain Adaptation for Clinical Negation Detection? *Journal of American Medical Informatics Association*. 2020.

Jian Zhang, Anil Can, Pui Man Rosalind Lai, Srinivasan Mukundan, Victor M Castro, Dmitriy Dligach, Sean Finan, Sheng Yu, Vivian S Gainer, Nancy A Shadick, Guergana Savova, Shawn N Murphy, Tianxi Cai, Scott T Weiss, Rose Du. Age and morphology of posterior communicating artery aneurysms. *Scientific Reports*. 2020.

Daniel To, Brihat Sharma, Niranjana Karnik, Cara Joyce, Dmitriy Dligach, Majid Afshar. Validation of an Alcohol Misuse Classifier in Hospitalized Patients. *Alcohol*. 2019.

Majid Afshar, Cara Joyce, Dmitriy Dligach, Brihat Sharma, Robert Kania, Meng Xie, Kristin Swope, Elizabeth Salisbury-Afshar, Niranjana Karnik. Subtypes in patients with opioid misuse: A prognostic enrichment strategy using electronic health record data in hospitalized patients. *PLoS One*. 2019.

Dmitriy Dligach, Majid Afshar, Timothy Miller. Towards a Clinical Text Encoder: Pre-training for Clinical Natural Language Processing with Applications to Substance Misuse. *Journal of American Medical Informatics Association*. 2019.

Majid Afshar, Dmitriy Dligach, Brihat Sharma, Xiaoyuan Cai, Jason Boyda, Steven Birch, Daniel Valdez, Suzan Zelisko, Cara Joyce, François Modave, Ron Price. Development and application of a high throughput natural language processing pipeline to convert all clinical documents in a clinical data warehouse into standardized medical vocabularies. *Journal of American Medical Informatics Association*. 2019.

Anil Can, Pui Man Rosalind Lai, Victor M Castro, Sheng Yu, Dmitriy Dligach, Sean Finan, Vivian Gainer, Nancy A Shadick, Guergana Savova, Shawn Murphy, Tianxi Cai, Scott T Weiss, Rose Du. Decreased Total Iron Binding Capacity May Correlate with Ruptured Intracranial Aneurysms. *Scientific Reports*. 2019.

Elena Rodina and Dmitriy Dligach. Dictator's Instagram: Personal and Political Narratives In a Chechen Leader's Social Network. *Caucasus Survey*. 2019.

Majid Afshar, Andrew Phillips, Anthony Oakey, Richard Cooper, Jeanne Meuller, Richard Gonzalez, Dmitry Dligach. Natural language processing and machine learning to identify alcohol misuse from the electronic health record in trauma patients: Development and internal validation. *Journal of American Medical Informatics Association*. 2019.

Anil Can, Victor M. Castro, Dmitriy Dligach, Sean Finan, Sheng Yu, Vivian Gainer, Nancy A. Shadick, Guergana Savova, Shawn Murphy, Tianxi Cai, Scott T. Weiss, and Rose Du. Elevated International Normalized Ratio is Associated with Ruptured Aneurysms. *Stroke*. 2018.

Anil Can, Robert F Rudy, Victor Castro, Sheng Yu, Dmitriy Dligach, Sean Finan, Vivian S Gainer, Nancy A Shadick, Guergana Savova, Shawn Murphy, Tianxi Cai, Scott Weiss, and Rose Du. Association between Aspirin Dose and Subarachnoid Hemorrhage from Saccular Aneurysms: A Case-Control Study. *Neurology*. 2018.

Anil Can, Victor M. Castro, Dmitriy Dligach, Sean Finan, Sheng Yu, Vivian Gainer, Nancy A. Shadick, Guergana Savova, Shawn Murphy, Tianxi Cai, Scott T. Weiss, Rose Du. Low Serum Calcium and Magnesium Levels and Rupture of Intracranial Aneurysms. *Stroke*. 2018.

Anil Can, Victor M Castro, Dmitriy Dligach, Sean Finan, Sheng Yu, Vivian Gainer, Nancy A Shadick, Guergana Savova, Shawn Murphy, Tianxi Cai, Scott T Weiss, Rose Du. Lipid-Lowering Agents and High HDL (High-Density Lipoprotein) Are Inversely Associated With Intracranial Aneurysm Rupture. *Stroke*. 2018.

Jian Zhang, Anil Can, Srinivasan Mukundan Jr., Michael Steigner, Victor M. Castro, Dmitriy Dligach, Sean Finan, Sheng Yu, Vivian Gainer, Nancy A. Shadick, Guergana Savova, Shawn Murphy, Tianxi Cai, Zhong Wang, Scott T. Weiss, Rose Du. Morphological Variables Associated with Ruptured Middle Cerebral Artery Aneurysms. *Neurosurgery*. 2018.

Anil Can, Victor M. Castro, Sheng Yu, Dmitriy Dligach, Sean Finan, Vivian Gainer, Nancy A. Shadick, Guergana Savova, Shawn Murphy, Tianxi Cai, Guergana Savova, Scott T. Weiss, Rose Du. Antihyperglycemic Agents are Inversely Associated with Intracranial Aneurysm Rupture. *Stroke*. 2017.

Anil Can, Victor M. Castro, Yildirim H. Ozdemir, Sarajune Dagen, Dmitriy Dligach, Sean Finan, Sheng Yu, Vivian Gainer, Nancy A. Shadick, Shawn Murphy, Tianxi Cai, Guergana Savova, Scott T. Weiss, Rose Du. Alcohol Consumption and Aneurysmal Subarachnoid Hemorrhage. *Translational Stroke Research*. 2017.

Anil Can, Victor M Castro, Yildirim H Ozdemir, Sarajune Dagen, Dmitriy Dligach, Sean Finan, Sheng Yu, Vivian Gainer, Nancy A Shadick, Guergana Savova, Shawn Murphy, Tianxi Cai, Scott T Weiss, Rose Du. Heroin use is associated with ruptured saccular aneurysms. *Translational Stroke Research*. 2017.

Anil Can, Victor Castro, Yildirim H Ozdemir, Sarajune Dagen, Sheng Yu, Dmitriy Dligach, Sean Finan, Vivian S Gainer, Nancy Shadick, Shawn Murphy, Tianxi Cai, Guergana Savova, Ruben Dammers, Scott T Weiss, and Rose Du. Association of Intracranial Aneurysm Rupture with Smoking Duration, Intensity, and Cessation. *Neurology*. 2017.

Timothy Miller, Dmitriy Dligach, Steven Bethard, Chen Lin, and Guergana Savovaa. Towards generalizable entity-centric clinical coreference resolution. *Journal of Biomedical Informatics*. 2017. doi: <https://doi.org/10.1016/j.jbi.2017.04.015>.

Victor M Castro, Dmitriy Dligach, Sean Finan, Sheng Yu, Anil Can, Muhammad Abd-El-Barr, Vivian Gainer, Nancy A Shadick, Shawn Murphy, Tianxi Cai, Guergana Savova, Scott T Weiss, and Rose Du. Large-scale Identification of Patients with Cerebral Aneurysms Using Natural Language Processing. *Neurology*. 2017. DOI: 10.1212/WNL.0000000000003490.

\*Chen Lin, \*Dmitriy Dligach, \*Timothy Miller, \*Steven Bethard, Guergana Savova. Multilayered Temporal Modeling for the Clinical Domain. *Journal of American Medical Informatics Association*. 2016. doi: 10.1093/jamia/ocv113. \* Indicates co-first authors

\*Chen Lin, \*Elizabeth W. Karlson, \*Dmitriy Dligach, Monica Ramirez, Timothy A. Miller, Huan Mo, Natalie S. Braggs, Andrew Cagan, Joshua C. Denny, Guergana K. Savova. Automatic Identification of Methotrexate-induced Liver Toxicity in Patients with Rheumatoid Arthritis from the Electronic Medical Record. *Journal of American Medical Informatics Association*. 2015. doi: 10.1136/amiajnl-2014-002642.

\* Indicates co-first authors

Dmitriy Dligach, Steven Bethard, Lee Becker, Timothy A. Miller, Guergana K. Savova. Discovering Body Site and Severity Modifiers in Clinical Texts. *Journal of American Medical Informatics Association*. 2014. 2014;21:438-447. doi:10.1136/amiajnl-2013-001949.

Chen Lin, Elizabeth W. Karlson, Helena Canhao, Timothy A. Miller, Dmitriy Dligach, Pei Jun Chen, Raul Natanael Guzman Perez, Tianxi Cai, Michael E. Weinblatt, Nancy A. Shadick, Robert M. Plenge, Guergana K. Savova. Automatic Prediction of Rheumatoid Arthritis Disease Activity from the Electronic Medical Records. *PLoS ONE*. 2013. 8(8): e69932. doi:10.1371/journal.pone.0069932.

Daniel Albright, Arrick Lanfranchi, Anwen Fredriksen, William F. Styler, Colin Warner, Jena D. Hwang, Jinho D. Choi, Dmitriy Dligach, Rodney D. Nielsen, James Martin, Wayne Ward, Martha Palmer, Guergana K. Savova. Towards comprehensive syntactic and semantic annotations of the clinical narrative. *Journal of American Medical Informatics Association*. 2013. Sep-Oct;20(5):922-30. doi: 10.1136/amiajnl-2012-001317.

Stephen T. Wu, Vinod C. Kaggal, Dmitriy Dligach, James J. Masanz, Pei Chen, Lee Becker, Wendy W. Chapman, Guergana K. Savova, Hongfang Liu and Christopher G. Chute. A common type system for clinical Natural Language Processing. *Journal of Biomedical Semantics*. January 2013, 4:1.

Dmitriy Dligach and Martha Palmer. Word Sense Disambiguation with Automatically Retrieved Semantic Knowledge. *International Journal of Semantic Computing (IJSC)*. 2:3, pp. 365-380, 2008.

## Conference Proceedings

Croxford, Emma Leigh, Yanjun Gao, Brian W. Patterson, Daniel Chern-Hui To, Samuel Tesch, Dmitriy Dligach, Anoop Mayampurath, Matthew M. Churpek, and Majid Afshar. Development of a Human Evaluation Framework and Correlation with Automated Metrics for Natural Language Generation of Medical Diagnoses. In *Proceedings of American Medical Informatics Association Symposium 2024*. San Francisco, CA. April 9, 2024.

Graham J Wills, Majid Afshar, MD, Brian Patterson, MD, Matthew Churpek MD, Dmitriy Dligach, Benjamin Schnapp, Michael Chao, Frank Liao. An Operational and Real-Time Natural Language Processing and Deep Learning Clinical Decision Support System in the Electronic Health Record. In *Proceedings of American Medical Informatics Association Symposium 2023*. New Orleans, LA. November, 2023.

Yanjun Gao, Dmitriy Dligach, Timothy A. Miller, Dongfang Xu, Matthew M. Churpek and Majid Afshar. Summarizing Patients Problems from Hospital Progress Notes Using Pre-trained Sequence-to-Sequence Models. In *Proceedings of the 29th International Conference On Computational Linguistics (COLING)*. Gyeongju. Republic of Korea. October, 2022.

Gao, Yanjun, Dmitriy Dligach, Timothy Miller, Samuel Tesch, Ryan Laffin, Matthew M. Churpek, and Majid Afshar. Hierarchical Annotation for Building A Suite of Clinical Natural Language Processing Tasks: Progress Note Understanding. In *Proceedings of the 13th Language Resources and Evaluation (LREC) Conference*. Marseille, France. June, 2022.

Daniel To, Cara Joyce, Sujay Kulshrestha, Brihat Sharma, Dmitriy Dligach, Mathew Churpek, Majid Afshar. The Addition of United States Census-Tract Data Does Not Improve the Prediction of Substance Misuse. In *Proceedings of American Medical Informatics Association Symposium 2021*. San Diego, CA. November, 2021.

Majid Afshar, Cara Joyce, Anthony Oakey, Perry Formanek, Philip Yang, Matthew M. Churpek, Richard S. Cooper, Susan Zelisko, Ron Price, Dmitriy Dligach. A Computable Phenotype for Acute Respiratory Distress Syndrome Using Natural Language Processing and Machine Learning. In *Proceedings of American Medical Informatics Association Symposium 2018*. San Francisco, CA. November, 2018.

Dmitriy Dligach and Timothy Miller. Learning Patient Representations from Text. In *\*SEM 2018: Proceedings of the 7th Joint Conference on Lexical and Computational Semantics*. New Orleans, LA. June, 2018.

Natalia Viani, Timothy Miller, Dmitriy Dligach, Steven Bethard, Carlo Napolitano, Silvia Priori, Riccardo Bellazzi, Lucia Sacchi and Guergana Savova. Recurrent Neural Network Architectures for Event Extraction from Italian Medical Reports. In *AIME 2017: Proceedings of the 16th Conference on Artificial Intelligence in Medicine*. Vienna, Austria. June, 2017.

Dmitriy Dligach, Timothy Miller, Chen Lin, Steven Bethard, and Guergana Savova. Neural Temporal Relation Extraction. In *EACL 2017: Proceedings of the 15th Meeting of the European Chapter of the Association for Computational Linguistics*. Valencia, Spain. April, 2017.

Dmitriy Dligach, Timothy Miller, Guergana Savova. Semi-supervised Learning for Phenotyping Tasks. In *Proceedings of American Medical Informatics Association Symposium 2015*. San Francisco, CA. November, 2015.

Chen Lin, Timothy Miller, Alvin Kho, Steven Bethard, Dmitriy Dligach, Sameer Pradhan and Guergana Savova. Descending-Path Convolution Kernel for Syntactic Structures. In *ACL '14: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. June 22 - 27, 2014, Baltimore, MD.

Dmitriy Dligach and Martha Palmer. Good Seed Makes a Good Crop: Accelerating Active Learning Using Language Modeling. In *ACL '11: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. June 19 - 24, 2011, Portland, OR.

Susan Windisch Brown, Dmitriy Dligach, and Martha Palmer. VerbNet Class Assignment as a WSD Task. In *IWSC 2011: Proceedings of the 9th International Conference on Computational Semantics*. January 12 - 14, 2011, Oxford, UK.

Dmitriy Dligach and Martha Palmer. Improving Verb Sense Disambiguation with Automatically Retrieved Semantic Knowledge. In *ICSC '08: Proceedings of the Second IEEE International Conference on Semantic Computing*. August 4 - 7, 2008, Santa Clara, CA.

Dmitriy Dligach and Martha Palmer. Novel Semantic Features for Verb Sense Disambiguation. In *ACL '08: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. June 15 - 20, 2008, Columbus, OH.

Jinying Chen, Dmitriy Dligach, and Martha Palmer. Towards Large-scale High-Performance English Verb Sense Disambiguation by Using Linguistically Motivated Features. In *ICSC '07: Proceedings of the First IEEE International Conference on Semantic Computing*. Sep 17 - 19, 2007, Irvine, CA.

Cecily Jill Duffield, Jena D. Hwang, Susan Windisch Brown, Dmitriy Dligach, Sarah E. Vieweg, Jenny Davis, and Martha Palmer. Criteria for the Manual Clustering of Verb Senses. In *CogSci '07: Proceedings of the 29th Annual Conference of the Cognitive Science Society*. August 1-4, 2007, Nashville, TN.

## Workshop Proceedings

YanJun Gao, Skatje Myers, Shan Chen, Dmitriy Dligach, Timothy A Miller, Danielle Bitterman, Guanhua Chen, Anoop Mayampurath, Matthew Churpek, Majid Afshar. Position Paper On Diagnostic Uncertainty Estimation from Large Language Models: Next-Word Probability Is Not Pre-test Probability. In *NeurIPS 2024 Workshop GenAI4Health Submission*. December 2024. Vancouver, Canada.

Brihat Sharma, YanJun Gao, Timothy Miller, Matthew Churpek, Majid Afshar, and Dmitriy Dligach. Multi-Task Training with In-Domain Language Models for Diagnostic Reasoning. In *Proceedings of The 5th Clinical Natural Language Processing Workshop (ClinicalNLP 2023) held in conjunction with ACL 2023*. July 2023. Toronto, Canada.

Weipeng Zhou, Majid Afshar, Dmitriy Dligach, YanJun Gao, and Timothy Miller. Improving the Transferability of Clinical Note Section Classification Models with BERT and Large Language Model Ensembles. In *Proceedings of The 5th Clinical Natural Language Processing Workshop (ClinicalNLP 2023) held in conjunction with ACL 2023*. July 2023. Toronto, Canada.

YanJun Gao, Dmitriy Dligach, Timothy Miller and Majid Afshar. Overview of the Problem List Summarization (ProbSum) 2023 Shared Task on Summarizing Patients' Active Diagnoses and Problems from Electronic Health Record Progress Notes. In *Proceedings of the 22nd BioNLP Workshop held in conjunction with ACL 2023*. July 2023. Toronto, Canada.

Timothy Miller, Steven Bethard, Dmitriy Dligach and Guergana Savova. End-to-end clinical temporal information extraction with multi-head attention. In *Proceedings of the 22nd BioNLP Workshop held in conjunction with ACL 2023*. July 2023. Toronto, Canada.

Dmitriy Dligach, Steven Bethard, Timothy Miller, Guergana Savova. Exploring Text Representations for Generative Temporal Relation Extraction. In *Proceedings of the 4th Clinical Natural Language Processing Workshop (cNLP 2022) held in conjunction with NAACL 2022*. July 2022.

Chen Lin, Timothy Miller, Dmitriy Dligach, Steven Bethard and Guergana Savova. EntityBERT: Entity-centric Masking Strategy for Model Pretraining for the Clinical Domain. In *Proceedings of the 2021 Workshop on Biomedical Natural Language Processing (BioNLP 2021) held in conjunction with NAACL 2021*. June 2021.

Kristin Wright-Bettner, Chen Lin, Timothy Miller, Steven Bethard, Dmitriy Dligach, Martha Palmer, James H. Martin and Guergana Savova. Defining and Learning Refined Temporal Relations in the Clinical Narrative. In *11th International Workshop on Health Text Mining and Information Analysis (LOUHI 2020) colocated with EMNLP 2020*. November 2020.

Chen Lin, Timothy Miller, Dmitriy Dligach, Farig Sadeque, Steven Bethard and Guergana Savova. A BERT-based One-Pass Multi-Task Model for Clinical Temporal Relation Extraction. In *Proceedings of the 2020 Workshop on Biomedical Natural Language Processing (BioNLP 2020) held in conjunction with ACL 2020*. July 2020.

Dianbo Liu, Dmitriy Dligach, and Timothy Miller. Two-stage Federated Phenotyping and Patient Representation Learning. In *Proceedings of the 2019 Workshop on Biomedical Natural Language Processing (BioNLP 2019) held in conjunction with ACL 2019*. August 2019. Florence, Italy.

Chen Lin, Timothy Miller, Dmitriy Dligach, Steven Bethard and Guergana Savova. A BERT-based Universal Model for Both Within- and Cross-sentence Clinical Temporal Relation Extraction. In *Proceedings of ClinicalNLP 2019: The 2nd Clinical Natural Language Processing Workshop held in conjunction with NAACL 2019*. June 2019. Minneapolis, MN.



Timothy Miller, Alon Geva and Dmitriy Dligach. Extracting Adverse Drug Event Information with Minimal Engineering. In *Proceedings of ClinicalNLP 2019: The 2nd Clinical Natural Language Processing Workshop held in conjunction with NAACL 2019*. June 2019. Minneapolis, MN.

Chen Lin, Timothy Miller, Dmitriy Dligach, Hadi Amiri, Steven Bethard and Guergana Savova. Self-training improves Recurrent Neural Networks performance for Temporal Relation Extraction. In *Proceedings of LOUHI 2018: The Ninth International Workshop on Health Text Mining and Information Analysis held in conjunction with EMNLP 2018*. November 2018. Brussels, Belgium.

Chen Lin, Timothy Miller, Dmitriy Dligach, Steven Bethard and Guergana Savova. Representations of Time Expressions for Temporal Relation Extraction with Convolutional Neural Networks. In *Proceedings of the 2017 Workshop on Biomedical Natural Language Processing (BioNLP 2017) held in conjunction with ACL 2017*. August 2017. Vancouver, Canada.

Timothy Miller, Dmitriy Dligach and Guergana Savova. Unsupervised Document Classification with Informed Topic Models. In *Proceedings of the 2016 Workshop on Biomedical Natural Language Processing (BioNLP 2016) held in conjunction with ACL 2016*. August 2016. Berlin, Germany.

Chen Lin, Timothy Miller, Dmitriy Dligach, Steven Bethard and Guergana Savova. Improving Temporal Relation Extraction with Training Instance Augmentation. In *Proceedings of the 2016 Workshop on Biomedical Natural Language Processing (BioNLP 2016) held in conjunction with ACL 2016*. August 2016. Berlin, Germany.

Timothy A. Miller, Steven Bethard, Dmitriy Dligach, Chen Lin, Guergana K. Savova. Extracting Time Expressions from Clinical Text. In *Proceedings of the 2015 Workshop on Biomedical Natural Language Processing (BioNLP 2015) held in conjunction with ACL 2015*. July 2015, Beijing, China.

Dmitriy Dligach, Timothy A. Miller, Guergana K. Savova. Active Learning for Phenotyping Tasks. In *Proceedings of the 2013 NLP for Medicine and Biology workshop held in conjunction with RANLP-2013*. September 2013. Hissar, Bulgaria.

Timothy A. Miller, Steven Bethard, Dmitriy Dligach, Sameer Pradhan, Chen Lin, Guergana K. Savova. Discovering Narrative Containers in Clinical Text. In *Proceedings of the 2013 Workshop on Biomedical Natural Language Processing (BioNLP 2013) held in conjunction with NAACL-2013*. August 2013, Sophia, Bulgaria.

Chen Lin, Timothy Miller, Dmitriy Dligach, Robert M. Plenge, Elizabeth W. Karlson, Guergana Savova. Maximal Information Coefficient for Feature Selection for Clinical Document Classification. In *International Conference on Machine Learning (ICML 2012) Workshop on Machine Learning for Clinical Data Analysis*. June 2012, Edinburgh, Scotland.

Chen Lin, Timothy Miller, Dmitry Dligach, Robert Plenge, Elizabeth Karlson, Guergana Savova. Feature Engineering and Selection for Rheumatoid Arthritis Disease Activity Classification Using Electronic Medical Records. In *International Conference on Machine Learning (ICML 2012) Workshop on Machine Learning for Clinical Data Analysis*. June 2012, Edinburgh, Scotland.

Timothy A. Miller, Dmitriy Dligach, and Guergana K. Savova. Active Learning for Coreference Resolution. In *Proceedings of the 2012 Workshop on Biomedical Natural Language Processing (BioNLP 2012) held in conjunction with NAACL-2012*. June 2012, Montreal, Canada.

Stephen T. Wu, Vinod C. Kaggal, Guergana K. Savova, Hongfang Liu, Dmitriy Dligach, Jiaping Zheng, Wendy W. Chapman, Christopher G. Chute. Generality and Reuse in a Common Type System for Clinical Natural Language Processing. In *Managing Interoperability and complexity in Health Systems (MIXHS-2011)*. October 2011, Glasgow, Scotland, U.K.

Dmitriy Dligach and Martha Palmer. Reducing the Need for Double Annotation. In *Proceedings of the Fifth Linguistic Annotation Workshop (LAW V) held in conjunction with ACL-2011*. June, 2011, Portland, OR.

Dmitriy Dligach, Rodney D. Nielsen, and Martha Palmer. To Annotate More or to Annotate More Accurately. In *Proceedings of the Fourth Linguistic Annotation Workshop (LAW IV) held in conjunction with ACL-2010*. July, 2010, Uppsala, Sweden.

Suresh Manandhar, Ioannis P. Klapaftis, Dmitriy Dligach, and Sameer S. Pradhan. SemEval-2010 Task 14: Word Sense Induction & Disambiguation. In *Proceedings of the Fifth International Workshop on Semantic Evaluations (SemEval-2010) held in conjunction with ACL-2010*. July, 2010, Uppsala, Sweden.

Dmitriy Dligach and Martha Palmer. Using Language Modeling to Select Useful Annotation Data. In *Proceedings of the Student Research Workshop and Doctoral Consortium Held in Conjunction with NAACL-HLT 2009*. May 31 - June 5, 2009. Boulder, Colorado.

Sameer Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. SemEval-2007 Task-17: English Lexical Sample, SRL and All Words. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007) held in conjunction with ACL-2007*. June, 2007, Prague, the Czech Republic.

Cecily Jill Duffield, Jena D. Hwang, Susan Windisch Brown, Dmitriy Dligach, Sarah E. Vieweg, Jenny Davis, and Martha Palmer. 2007. Criteria for the Manual Grouping of Verb Senses. In *Proceedings of the Linguistic Annotation Workshop held in conjunction with ACL 2007*. June, 2007, Prague, the Czech Republic.

## Book Chapters

Susan Windisch Brown, Dmitriy Dligach, and Martha Palmer. VerbNet Class Assignment as a WSD Task. 2014. *Computing Meaning Vol. 4*. Springer. pp 203-216.

## Conference Abstracts, Presentations, and Panels

Anoop Mayampurath, Abigail Hahs, John Caskey, Askar Safipour Afshar, Matthew Churpek, Dmitriy Dligach, Majid Afshar. Association Between Stigmatizing Language Occurrence in Notes and Outcomes in Patients with Sepsis. *Society of Critical Care Medicine's (SCCM) Critical Care Congress*. Orlando, FL. February 2025.

Behnaz Eslami, Benjamin Strickland, Holly K. Kramer, Dmitriy Dligach, Samie Tootooni. A Performance-Based Voting Framework for Structuring Clinical Notes. *American Medical Informatics Association Annual Symposium*. San Francisco, CA. November 2024.

Jifan Gao, Guanhua Chen, Ann P O'Rourke, John Caskey, Kyle A Carey, Madeline Oguss, Anne Stey, Dmitriy Dligach, Timothy Miller, Anoop Mayampurath, Matthew M Churpek, Majid Afshar. Automated Stratification of Trauma Injury Severity Across Multiple Body Regions Using Multimodal, Multiclass Machine Learning Models *American Medical Informatics Association Annual Symposium*. San Francisco, CA. November 2024.

Brandon Leding, Yanjun Gao, Dmitriy Dligach, Emma Croxford, Anoop Mayampurath, Mathew Churpek, Majid Afshar. Generation and Evaluation of Synthetic Critical Care Progress Notes With Large Language Models. *American Thoracic Society International Conference*. San Diego, CA. 2024.

Shan Chen, Marco Guevara, Yanjun Gao, Majid Afshar, Timothy Miller, Dmitriy Dligach, Danielle S. Bitterman. Improving Clinical NLP Performance through Language Model-Generated Synthetic Clinical Data. *American Medical Informatics Association Summit*. Boston, MA. March 2024.

Anoop Mayampurath, Kyle Carey, Priti Jani, Majid Afshar, Matthew Churpek, and Dmitry Dligach. Explaining Alerts from a Pediatric Deterioration Prediction Model Using Clinical Text. *American Medical Informatics Association Annual Symposium*. Washington, DC. November 2022. **AMIA 2022 Distinguished poster award.**

John Caskey, Iain L. McConnell, Madeline Oguss, Dmitriy Dligach, Rachel Kulikoff, Brittany Grogan, Crystal Gibson, Elizabeth Wimmer, Traci E DeSalvo, Edwin E Nyakoe-Nyasani, Matthew M Churpek, Majid Afshar. *Pragmatic Design and Application of a Named Entity Recognition Pipeline to Assist Contact Tracers during the COVID-19 Pandemic*. AMIA Informatics Summit. Chicago, IL. March 2022.

Brihat Sharma, Dmitriy Dligach, Hale M. Thomson, Matthew M Churpek, Niranjana S. Karnik, Majid Afshar. A multi-label classifier to screen different types of substance misuse in hospitalized patients. *American Medical Informatics Association Annual Symposium*. San-Diego, CA. November 2021.

Sujay Kulshrestha, Dmitriy Dligach, Xin Su, Richard Gonzalez, Cara Joyce, Matthew M Churpek, Majid Afshar. Classification of Chest Injury Severity Using Clinical Documents. AMIA Informatics Summit. 2021.

Dmitriy Dligach, Sean Finan, Peter Abramowitsch. Apache cTAKES and Python; Apache cTAKES High Throughput Orchestration. *ApacheCon*. October 2020.

Timothy Miller, Dmitriy Dligach, Fei Wang, Yuqi Si. Learning Patient Representations from Electronic Health Records across Diverse Data Types. Panel. *American Medical Informatics Association Symposium*. Chicago, IL. November 2020.

Xin Su, Timothy Miller, Majid Afshar, Dmitriy Dligach. Learning Hierarchical Transformer-based Representations of Clinical Notes. *American Medical Informatics Association Symposium*. Chicago, IL. November 2020.

Sujay Kulshrestha, Brihat Sharma, Richard Gonzalez, Cara Joyce, Dmitriy Dligach, Matthew M. Churpek, Majid Afshar. Use of Machine Learning to Predict Severity of Chest Injury from Clinical Texts. *AMIA 2020 Clinical Informatics Conference*. Seattle, WA. May 2020.

Xin Su, Anoop Mayampurath, Matthew Churpek, Sameep Shah, Bhakti Patel, Dmitriy Dligach, and Majid Afshar. External Validation of an Acute Respiratory Distress Syndrome Prediction Model Using Clinical Text. *American Thoracic Society (ATS) International Conference 2020*. Philadelphia, Pennsylvania, May 2020.

Daniel To, Brihat Sharma, Cara Joyce, Dmitriy Dligach, Elizabeth Salisbury-Afshar, Niranjana Karnik, Majid Afshar. Subtypes of Alcohol Misuse in Hospitalized Patients. *AMIA 2020 Informatics Summit*. Houston, TX, March 2019.

Xin Su, Timothy Miller, Farig Sadeque, Majid Afshar, Dmitriy Dligach. Using Transformer-based Approaches for Measuring Semantic Similarity. *N2C2 2019 Shared Task Workshop at AMIA 2019*. Washington DC. November, 2019.

Brihat Sharma, Dmitriy Dligach, Cara Joyce, Robert Kania, Elizabeth Salisbury-Afshar, Niranjana Karnik, Majid Afshar. Identification of Latent Subtypes of Patients with Opioid Misuse. *American Medical Informatics Association Symposium*. Washington, DC. November 2019.

Amy L. Olex, Tamas Gal, Majid Afshar, Dmitriy Dligach, Niranjana Karnik, Travis Oakes, Brihat Sharma, Meng Xie, Bridget T. McInnes, Julian Solway, Abel Kho, William C. Cramer, and F. Gerard Moeller. Untapped Potential of Clinical Text for Opioid Surveillance. *American Medical Informatics Association Symposium*. Washington, DC. November 2019.

Dmitriy Dligach, Majid Afshar, Timothy Miller. Towards a Universal Document-Level Clinical Text Encoder: Methods for Neural Network Pre-training with Applications to Substance Misuse. *American Medical Informatics Association Symposium*. Washington DC, November, 2019.

Daniel To, Andrew Phillips, Meng Xie, Dmitriy Dligach, Richard Gonzalez, Niranjin Karnik, Cara Joyce, Majid Afshar, "Development and Validation of a Computable Phenotype for Identification of Alcohol Misuse in Hospitalized Patients". Poster presentation at the 42nd Annual Research Society of Alcoholism (RSA) Scientific Meeting, Minneapolis, Minnesota. June 22-26 2019.

Brihat Sharma, Dmitriy Dligach, Cara Joyce, Stephen Littleton, and Majid Afshar. Phenotypic Clusters Derived from Clinical Notes of Patients with Respiratory Failure. The American Thoracic Society (ATS) 2019. Dallas, TX. May 2019.

Dmitriy Dligach and Timothy A. Miller. Loyola and Boston Children's Hospital at N2C2 Challenge Track 1: Cohort Selection for Clinical Trials. N2C2 2018 Shared Task Workshop at AMIA 2018. San-Francisco, CA. November, 2018.

Timothy A. Miller and Dmitriy Dligach. Extracting Drug Information with Apache cTAKES and ClearTK. N2C2 2018 Shared Task Workshop at AMIA 2018. San-Francisco, CA. November, 2018.

Anthony Oakey, Dmitry Dligach, Philip Yang, Perry Formanek, Susan Zelisko, Ron Price, Cara Joyce, Richard Cooper, Majid Afshar. Natural language processing and machine learning for identification of Acute Respiratory Distress Syndrome. The American Thoracic Society (ATS) 2018. San Diego, CA. May, 2018.

Dmitriy Dligach, Timothy Miller, Chen Lin, Steven Bethard, and Guergana Savova. Neural Models for Clinical Temporal Relation Extraction. Midwest Speech and Language Days & Midwest Computational Linguistics Colloquium 2017. Chicago, IL. May, 2017.

Timothy A. Miller, Dmitriy Dligach, Chen Lin, Steven Bethard, Guergana K. Savova. Cross-domain Coreference Feature Exploration. AMIA Annual Symposium. Chicago, IL. November, 2016.

Elena Rodina and Dmitriy Dligach. An on-line dictator; A Textual Analysis of the Instagram Account of Ramzan Kadyrov, President of the Chechen Republic. Amsterdam Text Analysis Conference. Amsterdam, Netherlands. June 2016.

Timothy A. Miller, Sean Finan, Dmitriy Dligach, Guergana Savova. Robust Sentence Segmentation for Clinical Text. AMIA Annual Symposium. San-Francisco, CA. November, 2015.

Dmitriy Dligach, Steven Bethard, Lee Becker, Timothy A. Miller, Guergana K. Savova. Discovering Body Site and Severity Modifiers in Clinical Texts. In *American Medical Informatics Association Symposium*. Washington DC, November, 2013.

Chen Lin, Elizabeth W. Karlson, Helena Canhao, Timothy A. Miller, Dmitriy Dligach, Pei Jun Chen, Raul Natanael Guzman Perez, Tianxi Cai, Michael E. Weinblatt, Nancy A. Shadick, Robert M. Plenge, Guergana K. Savova. Automatic Prediction of Rheumatoid Arthritis Disease Activity from the Electronic Medical Records. In *American Medical Informatics Association Symposium*. Washington DC, November, 2013.

Timothy Miller, Dmitriy Dligach, Steven Bethard, Sameer Pradhan, Chen Lin, Guergana K. Savova. Discovering Time Expressions in Clinical Text. In *American Medical Informatics Association Symposium*. Washington DC, November, 2013.

Susan Windisch Brown, Cecily Jill Duffield, Jena D. Hwang, Dmitriy Dligach, Sarah E. Vieweg, Jenny Davis, Martha Palmer. The role of functional nouns in manual grouping of verb senses. In *Poster Session of Concept Types and Frames in Language, Cognition, and Science*. August 20-22, 2007, Dusseldorf, Germany.

## *Invited Talks*

Automatic Phenotyping in the Age of Deep Learning. DePaul University Research Colloquium. September 2021.

Pretraining phenotyping classifiers. University of Wisconsin Madison. Churpek/Afshar lab. December 2020.

Pretraining NLP classifiers. University of Chicago, Matt Churpek lab. February 2018.

Clinical Natural Language Processing with cTAKES. Loyola University Stritch School of Medicine. Chicago, IL. August, 2018.

Neural Models for Temporal Relation Extraction. Northwestern University, Department of Computer Science. Chicago, IL. May 2017.

Semantic Processing of Clinical Texts. Loyola University Stritch School of Medicine. Chicago, IL. March 2016.

Semantic Analysis of Clinical Texts. Northwestern University Feinberg School of Medicine. Chicago, IL. June 2015.

Semantic Analysis of Clinical Narratives. NorthShore University Health System Research Institute. Evanston, IL. June 2014.

Active Learning for Phenotyping Tasks. i2b2 3rd Annual Academic User Group Conference and NLP Workshop. Boston, MA. June 2013.

Discovering Body Site and Severity Modifiers in Rheumatoid Arthritis Data. i2b2 2nd Annual Academic User Group Conference and NLP Workshop. Boston, MA. June 2012.

Discovering Severity and Body Site Modifiers: A Relation Extraction Task. Strategic Health IT Advanced Research Project Annual Meeting. Rochester, MN. June, 2012.

## *Shared Tasks*

BioNLP Shared Task. Organizer. 2023.

National NLP Clinical Challenges (n2c2) Shared Task. Organizer. 2022.

## *Datasets and Resources*

Yanjun Gao, Timothy Miller, Majid Afshar, and Dmitriy Dligach. BioNLP Workshop 2023 Shared Task 1A: Problem List Summarization (version 1.0.0). Dataset. PhysioNet. 2023. <https://doi.org/10.13026/1z6g-ex18>.

Yanjun Gao, John Caskey, Timothy Miller, Brihat Sharma, Matthew Churpek, Dmitriy Dligach, Majid Afshar. Tasks 1 and 3 from Progress Note Understanding Suite of Tasks: SOAP Note Tagging and Problem List Summarization (version 1.0.0). Dataset. PhysioNet. 2022. <https://doi.org/10.13026/wkso-w041>.

Chen Lin, Steven Bethard, Guergana Savova, Timothy Miller, Dmitriy Dligach. EntityBERT: BERT-based Models Pretrained on MIMIC-III with or without Entity-centric Masking Strategy for the Clinical Domain (version 1.0.1). PhysioNet. 2022. <https://doi.org/10.13026/e7kt-q579>.

### *Ph.D. Thesis*

Dmitriy Dligach. High-performance Word Sense Disambiguation with Less Manual Effort. Ph.D. thesis, University of Colorado at Boulder, August 2010.

### *Minor Contributions*

William J. Rapaport. Philosophy of Computer Science: An Introductory Course. In *Teaching Philosophy* 28(4): 319-341. 2005.

### *Courses Taught*

COMP 379/479: Machine Learning. Loyola University Chicago. Fall 2016-2023.

COMP 329/429: Natural Language Processing. Loyola University Chicago. Spring 2018-2023.

COMP 170: Introduction to Object-Oriented Programming. Loyola University Chicago. 2016-2017.

CSCI 3202: Introduction to Artificial Intelligence. University of Colorado, Boulder, Invited lecture, Fall 2008.

LING 7800: Advanced Computational Linguistics: Lexical Semantics. University of Colorado, Boulder, Invited lecture, Spring 2007.

LING 5200: Computational Corpus Linguistics. University of Colorado, Boulder, Invited lecture, Spring 2006

CSE/PHI 510: Philosophy of Computer Science. State University of New York, Buffalo, Teaching Assistant, Fall 2004.

CSE 116 : Introduction To Computer Science for Majors II. State University of New York, Buffalo, Teaching Assistant, Fall 2003, Spring 2004.

### *Courses Prepared to Teach*

Natural Language Processing

Machine Learning

Deep Learning

Artificial Intelligence

Information Retrieval

Data Mining

### *Service*

BioNLP Workshop at ACL 2024. Program committee and paper reviewer. 2024.

ClinicalNLP Workshop at NAACL 2024. Program Committee and paper reviewer. 2024.

ClinicalNLP Workshop at ACL 2023. Paper Reviewer. 2023.

BioNLP Workshop at ACL 2023. Paper Reviewer. 2023.

Association for Computational Linguistics (ACL) conference. Paper Reviewer. 2023.

Journal of Biomedical Informatics. Paper reviewer. 2023.

2023 BioNLP Shared Task and Workshop. Paper reviewer. 2023.

2022 n2c2 Shared Task and Workshop. Paper reviewer. 2022.

Conference on Empirical Methods in Natural Language Processing (EMNLP). Paper Reviewer. 2022.

Clinical NLP (cNLP) Workshop at NAACL. Paper Reviewer. 2022.

Biomedical NLP (BioNLP) Workshop. Paper Reviewer. 2022.

Journal of Biomedical Informatics. Paper Reviewer. October 2020.

Journal of American Medical Informatics Association. Paper Reviewer. September 2020.

Clinical NLP Workshop at EMNLP. Program Committee. 2020.

Clinical NLP Workshop at EMNLP. Paper Reviewer. 2020.

Association for Computational Linguistics (ACL) conference. Paper Reviewer. 2020.

Journal of American Medical Informatics Association. Paper Reviewer. April 2019.

Association for Computational Linguistics (ACL) conference. Program Committee. 2019.

Association for Computational Linguistics (ACL) conference. Paper Reviewer. 2019.

North American Chapter of the Association for Computational Linguistics (NAACL) conference. Program Committee. 2019.

North American Chapter of the Association for Computational Linguistics (NAACL) conference. Paper Reviewer. 2019.

Clinical NLP Workshop at COLING. Program Committee. 2019.

Clinical NLP Workshop at COLING. Paper Reviewer. 2019.

North American Chapter of the Association for Computational Linguistics (NAACL) conference. Program Committee. December 2018.

North American Chapter of the Association for Computational Linguistics (NAACL) conference. Paper Reviewer. December 2018.

International Conference on Computational Linguistics (COLING). Paper Reviewer. April 2018.

Association for Computational Linguistics (ACL) conference. Program Committee. March 2018.

Association for Computational Linguistics (ACL) conference. Paper Reviewer. February 2018.

American Medical Informatics Association (AMIA) Symposium. Paper Reviewer. March 2017.

Association for Computational Linguistics (ACL) conference. Paper Reviewer. February 2017.

Clinical NLP Workshop at COLING 2016. Program Committee. December 2016.

Clinical NLP Workshop at COLING 2016. Paper Reviewer. December 2016.

American Medical Informatics Association (AMIA) Symposium. Paper Reviewer. April 2016.

Association for Computational Linguistics (ACL) conference. Paper Reviewer. March 2016.

American Medical Informatics Association (AMIA) Symposium. Paper Reviewer. April 2015.

Association for Computational Linguistics (ACL) conference. Program Committee. February 2015.

Association for Computational Linguistics (ACL) conference. Paper Reviewer. February 2015.

SemEval 2015. Paper Reviewer. February 2015.

International Conference on Computational Semantics (IWCS). Program Committee. January 2015.

International Conference on Computational Semantics (IWCS). Paper Reviewer. January 2015.

RANLP NLP for Medicine and Biology Workshop. Paper Reviewer. September 2013.

American Medical Informatics Association (AMIA) Symposium. Paper Reviewer. May 2013.

International Conference on Computational Semantics (IWCS). Program Committee. March 2013.

International Conference on Computational Semantics (IWCS). Paper Reviewer. March 2013.

i2b2/VA/Cincinnati 2011 Shared Task and Workshop. Paper Reviewer. September 2011.

Computational Linguistics. Paper Reviewer. May 2011.

SemEval 2010. Program Committee. April 2010.

SemEval 2010. Paper Reviewer. April 2010.

SemEval 2010 Tasks 14: Word Sense Induction & Disambiguation Task. Organizing Committee. April 2010.

### *Mentoring*

Scholarship Oversight Committee. Jiarui Yao. Boston Children's Hospital. Computational Health Informatics Program. 2022-2023.

Carbon fellowship mentor. Laura Maskeri. 2020.

K award mentor. Anoop Mayampurath. University of Chicago. 2020.

MS thesis advisor. Robert Kania. Predicting Drug Misuse Status Using Machine Learning on Electronic Health Records. 2020.

MS thesis reader. Paul Okoro. Optimizing Gene Expression Prediction and Omics Integration in Populations of African Ancestry.

MS thesis advisor. Brihat Sharma. Opioid misuse detection in hospitalized patients using convolutional neural networks. 2019.

MS thesis advisor. Andrew Phillips. Using Natural Language Processing and Machine Learning for the Identification of Alcohol Misuse in Trauma Patients. 2018.



## Funding

Ro1. NIH. A Novel Clinical Decision Support Tool for Predicting Deterioration in Hospitalized Children (). Site PI. Loyola budget: \$87,417 over 5 years.

Ro1. NIH/NLM. Learning Universal Patient Representations from Clinical Text with Hierarchical Transformers (Ro1LMo12973-04). Site PI. Loyola budget: \$325,318 over 4 years.

Ro1. NIH/NIDDK. Using Machine Learning for Early Recognition and Personalized Treatment of Acute Kidney Injury (1Ro1DK126933). Site PI. Loyola budget: \$82,640 over 5 years.

Ro1. NIH/NHLBI. Developing a clinical decision support tool for the identification, diagnosis, and treatment of critical illness in hospitalized patients (1Ro1HL157262). Site PI. Loyola budget: \$200,441 over 5 years.

Ro1. NIH/NIDA. Data Driven Strategies for Substance Misuse Identification in Hospitalized Patients (Ro1DA051464). Site PI. Loyola budget: \$491,973 over 5 years.

Ro1. NIH/NLM. Temporal relation discovery for clinical text (Ro1LMo10090). Site PI. Loyola budget: \$202,301 over 4 years. 2019.

Ro1. NIH/NLM. Learning Universal Patient Representations from Clinical Text with Hierarchical Recurrent Neural Networks (Ro1LMo12973). Site PI. Loyola budget: \$229,248 over 3 years. 2019.

NIDA clinical trial network. Great Lakes Node (UG1DA049467). Resource faculty.

Loyola University Chicago. Center for Health Outcomes and Informatics Research (CHOIR) grant. Identification of substance misuse in hospitalized patients: using readily available electronic health record data to build computable phenotypes for surveillance and screening. \$50,000 for 1 year. 2019.

Loyola University Chicago. Center for Health Outcomes and Informatics Research (CHOIR) grant. RA funding for two semesters.

'TSM 2.0: Advancing Translational Science in Metropolitan Chicago'. RA funding for two semesters.

Loyola University Chicago, Research Support Grant. Natural Language Processing for Phenotyping Tasks. PI: Dmitriy Dligach. May 2017. \$4,998.

NVIDIA Hardware Grant Program. Titan X GPU. July 2016. About \$1,200 value.

Boston Children's Hospital, The Program for Patient Safety and Quality. "Natural Language Processing for Quality Metrics". PI: Dmitriy Dligach. July 2015. \$19,388.

## Honors

Sujack Family Award for Excellence in Faculty Research. Loyola University Chicago. 2021.

Presidential Fellowship, State University of New York, Buffalo.

Fr. Gerst Memorial Award ("presented to a single graduating senior who has displayed outstanding academic achievement in mathematics or computer science"), Loyola University, Chicago, 1999.

Graduated Magna Cum Laude, Loyola University, Chicago, 1999.

Loyola University scholarship, Loyola University, Chicago, 1996.