

# Masoud Rouhizadeh MA, Prof.MS, MS, PhD

---

CONTACT	Malachowsky Hall for Data Science & IT University of Florida 1889 Museum Rd. Gainesville FL 32611	<i>email:</i> mrouhizadeh@ufl.edu <i>phone:</i> +1 352 273 9397 <i>website:</i> <a href="http://plaza.ufl.edu/mrouhizadeh">http://plaza.ufl.edu/mrouhizadeh</a>
RESEARCH INTERESTS	AI in Healthcare, Natural Language Processing; Machine Learning; Biomedical Informatics.	
EDUCATION & TRAINING	<b>Postdoctoral Training in Biomedical Informatics and Data Science</b> University of Pennsylvania	2015-2017
	<b>PhD in Computer Science &amp; Engineering</b> Oregon Health & Science University	2009-2015
	<b>MSc in Computer Science &amp; Engineering</b> Oregon Health & Science University	2009-2013
	<b>Professional Master in Human Language Technology &amp; Interfaces</b> University of Trento, Italy	2008-2009
	<b>MA in Linguistics</b> Allameh Tabatabai University	2004-2008
ACADEMIC EXPERIENCE	<b>University of Florida - Assistant Professor</b> Affiliations: <ul style="list-style-type: none"><li>• UF AI in the Health Sciences Initiative</li><li>• Center for Drug Evaluation and Safety (CoDES)</li><li>• Intelligent Critical Care Center (IC3)</li><li>• Dept of Pharmaceutical Outcomes and Policy — College of Pharmacy</li></ul> Research work includes applying AI methods for healthcare applications: <ul style="list-style-type: none"><li>• Clinical natural language processing</li><li>• Identifying social determinants of health</li><li>• Mental health assessment and support</li><li>• Substance use and opioid-related outcomes</li><li>• Analyzing cognitive and language impairments</li></ul>	2021-present
	<b>Johns Hopkins University - Adjunct Assistant Professor</b> Affiliations: <ul style="list-style-type: none"><li>• Biomedical Informatics and Data Science (BIDS)</li><li>• Center for Population Health IT (CPHIT)</li></ul> Research, funding, and educational collaborations in AI in healthcare:	2021-present

- Teaching graduate course: Natural Language Processing in the Health Sciences
- Leading projects on extracting social determinants of health from free text notes

**Johns Hopkins University - NLP Lead, Faculty Instructor**

2017-2021

Affiliations:

- Biomedical Informatics and Data Science (BIDS)
- Institute for Clinical and Translational Research (CTSA)
- Center for Population Health IT (CPHIT)
- Center for Language and Speech Processing (CLSP)
- Precision Medicine Analytics Platform (PMAP)

Led large-scale enterprise AI solutions for transnational applications in healthcare:

- Built and deployed deep learning NLP solutions on JHU Precision Medicine platform
- Extracted and standardized clinical concepts from unstructured EHRs into the OMOP CDM
- Standardized 350 million clinical notes and converted into interoperable HL7 format
- Developed NLP frameworks for assessing language impairment
- Built neural frameworks to extract social and behavioral factors from clinical notes
- Analyzed EHRs to understand mental health outcomes, suicide risk factors
- Led JHU COVID-19 NLP research collaborating with National COVID Cohort Collaborative

**University of Pennsylvania - Postdoctoral Research Scientist**

2015-2017

Affiliations:

- Department of Biostatistics and Informatics
- Department of Computer and Information Science
- Computer Science Department, Stony Brook University

Research and implementation projects in data science for clinical and mental health outcomes:

- Developed computational models to explore medication side effects and birth defects
- Performed empirically-driven verb clustering and dependency-based analysis of pronouns
- Built NLP pipeline for noisy text processing and feature extraction
- Collaborated on the Differential Language Analysis ToolKit (DLATK)

**Johns Hopkins University - Senior Affiliate**

Summer 2016

Affiliation:

- Center for Language and Speech Processing Frederick Jelinek Memorial Summer Workshop

Contribution:

- Identified mental health risk factors by linking social media data with electronic health records

**Oregon Health & Science University - Graduate Research Assistant** 2009-2015

Affiliation:

- Computer Science and Electrical Engineering
- Center for Spoken Language Understanding

Contributions:

- Developed distributional semantic models for characterization of impaired language (PhD)
- Designed and built VigNet - a frame-based ontology for text-to-scene conversion (MSc)

### **Columbia University - Visiting Scholar**

Summer 2012

Affiliation:

- Computer Science Department
- Spoken Language Processing Group

Contribution:

- Incorporated VigNet ontology into WordsEye text-to-scene system

### **University of Trento - Graduate Student**

2008-2009

Affiliations:

- Department of Information Engineering & Computer Science
- Centre for Mind/Brain sciences (CIMeC)

Contributions:

- Built a dialogue system for local restaurant search
- Conducted behavioral studies of lexical hierarchies in the mental lexicon

### **Fondazione Bruno Kessler - Research Intern**

Summer 2009

Affiliation:

- Human Language Technology Group

Contribution:

- Built a morphological analyzer for Persian using finite-state transducers

### **Shahid Beheshti University - Research Scientist**

2007-2009

Affiliation:

- Computer & Electrical Engineering Department
- Natural Language Processing Lab

Contribution:

- Developed FarsNet - the Persian WordNet
- Co-developed SBUQA question-answering system using Lexical Functional Grammar

## **PUBLICATIONS**

### **Journal Articles**

1. Elise Omaki, Wendy Shields, **Masoud Rouhizadeh**, Pamela Delgado-Barroso, Ruth Stefanos, Andrea Gielen. 2023. "Understanding the circumstances of paediatric fall injuries: a machine learning analysis of NEISS narratives". *Injury Prevention*.
2. Geoffrey Gray, Ayah Zirikly, Luis M. Ahumada, **Masoud Rouhizadeh**, Thomas Richards, Christopher Kitchen, Iman Foroughmand, and Elham Hatef. 2023. "Application of natural

language processing to identify social needs from patient medical notes: development and assessment of a scalable, performant, and rule-based model in an integrated healthcare delivery system.” *JAMIA open* 6, no. 4: ooad085.

3. S. Liu, A. Wen, L. Wang, H. He, S. Fu, R. Miller, A. Williams, D. Harris, R. Kavuluru, M. Liu, N. Abu-el-Rub, D. Schutte, R. Zhang, **Masoud Rouhizadeh**, J. D. Osborne, Y. He, U. Topaloglu, S. S. Hong, J. H. Saltz, T. Schaffter, E. Pfaff, C. G. Chute, T. Duong, M. A. Haendel, R. Fuentes, P. Szolovits, H. Xu, H. Liu. 2023. “An open natural language processing (NLP) framework for EHR-based clinical research: a case demonstration using the National COVID Cohort Collaborative (N3C)”. *Journal of the American Medical Informatics Association*.
4. Ali Zolnour, Christina E Eldredge, Anthony Faiola, Yadollah Yaghoobzadeh, Masoud Khani, Doreen Foy, Maxim Topaz, Hadi Kharrazi, Kin Wah Fung, Paul Fontelo, Anahita Davoudi, Azade Tabaie, Scott A Breitingner, Tyler S Oesterle, **Masoud Rouhizadeh**, Zahra Zonnor, Hans Moen, Timothy B Patrick, Maryam Zolnoori. 2023. “A risk identification model for detection of patients at risk of antidepressant discontinuation”. *Frontiers in Artificial Intelligence*.
5. Dawei Guan, Piaopiao Li, Vivian Fonseca, Lizheng Shi, Mohammed K Ali, Jithin Sam Varghese, Rodrigo M Carrillo-Larco, **Masoud Rouhizadeh**, Almut G Winterstein, Tianze Jiao, Hui Shao. 2023. “1005-P: Developing a Machine-Learning–Based Prediction Model for Diabetes Duration Using Information from Electronic Health Records”. *Journal of Diabetes*.
6. W. Huang, M. Ahmed, S. Smith, M. Hasan, **Masoud Rouhizadeh**, J. Bian, S. Kimmel, E. Morris, L. Yang, J. Guo. 2023. “EPH147 Trajectories of Sacubitril/Valsartan Adherence Among Medicare Beneficiaries with Heart Failure”. *Value in Health*.
7. Ruba Sajdeya, Mamoun T Mardini, Patrick J Tighe, Ronald L Ison, Chen Bai, Sebastian Jugl, Gao Hanzhi, Kimia Zandbiglari, Farzana I Adiba, Almut G Winterstein, Thomas A Pearson, Robert L Cook, **Masoud Rouhizadeh**. 2023. “Developing and validating a natural language processing algorithm to extract preoperative cannabis use status documentation from unstructured narrative clinical notes”. *Journal of the American Medical Informatics Association*.
8. V. K. Keloth, J. M. Banda, M. Gurley, P. M. Heider, G. Kennedy, H. Liu, F. Liu, T. Miller, K. Natarajan, O. V. Patterson, Y. Peng, K. Raja, R. M. Reeves, **Masoud Rouhizadeh**, J. Shi, X. Wang, Y. Wang, W.-Q. Wei, A. E. Williams, R. Zhang, R. Belenkaya, C. Reich, C. Blacketer, P. Ryan, G. Hripcsak, N. Elhadad, H. Xu. 2023. “Representing and Utilizing Clinical Textual Data for Real World Studies: An OHDSI Approach”. *Journal of Biomedical Informatics*.
9. Elham Hatef, **Masoud Rouhizadeh**, Claudia Nau, Fagen Xie, Christopher Rouillard, Mahmoud Abu-Nasser, Ariadna Padilla, Lindsay Joe Lyons, Hadi Kharrazi, Jonathan P Weiner, Douglas Roblin. 2022. “Development and assessment of a natural language processing model to identify residential instability in electronic health records’ unstructured data: a comparison of 3 integrated healthcare delivery systems”. *JAMIA Open*.
10. Jessica Schwartz, Eva Tseng, Nisa Maruthur, **Masoud Rouhizadeh**. 2022. “Identification of Prediabetes Discussions in Unstructured Clinical Documentation: Validation of a Natural Language Processing Algorithm”. *JMIR Medical Informatics*.
11. Elham Hatef, **Masoud Rouhizadeh**, Claudia Nau, Fagen Xie, Christopher Rouillard, Mahmoud Abu-Nasser, Ariadna Padilla, Lindsay Joe Lyons, Hadi Kharrazi, Jonathan P. Weiner, Douglas Roblin. 2022. “Development and assessment of a natural language processing model to identify residential instability in electronic health records’ unstructured data: a comparison of 3 integrated healthcare delivery systems”. *JAMIA Open*.
12. Golnoosh Alipour-Haris, Melissa J Armstrong, Jennifer Sullivan, Uma Suryadevara, **Masoud Rouhizadeh**, Joshua D Brown. 2022. “Suicidal Ideation and Suicide-Attempt-Related Hospitalizations among People with Alzheimer’s Disease (AD) and AD-Related Dementias in the United States during 2016–2018.” *Journal of Clinical Medicine*.
13. Elham Hatef, **Masoud Rouhizadeh**, Claudia Nau, Fagen Xie, Christopher Rouillard, Mahmoud Abu-Nasser, Douglas Roblin. 2021. “A Pilot Study to Improve the Use of Electronic

Health Records for Identification of Patients with Social Determinants of Health Challenges: A Collaboration of Johns Hopkins Health System and Kaiser Permanente' Unstructured Data: A Comparison of Three Integrated Healthcare Delivery Systems." *Health Services Research*.

14. Elham Hatef, Gurmehar Singh Deol, **Masoud Rouhizadeh**, Ashley Li, katyuska eibensteiner, Craig B. Monsen, Roman Bratslaver, Margaret Senese, Hadi Kharrazi. 2021. "Measuring the Value of Free-Text Notes in Electronic Health Record to Identify Patients with Housing Issues: Findings of a Retrospective Cohort Study." *Frontiers in Public Health*.
15. Jingqi Wang, Noor Abu-el-rub, Josh Gray, Huy Anh Pham, Yujia Zhou Frank Manion, Mei Liu, Xing Song, Hua Xu, Yaoyun Zhang, **Masoud Rouhizadeh**. 2021. "COVID-19 SignSym – A fast adaptation of a general clinical NLP tool to identify and normalize COVID-19 signs and symptoms to OMOP common data mode." *Journal of the American Medical Informatics Association*.
16. Eva Tseng, Jessica Schwartz, **Masoud Rouhizadeh**, Nisa Maruthur. 2021. "Analysis of Primary Care Provider Electronic Health Record Notes for Discussions of Prediabetes Using Natural Language Processing Methods." *Journal of General Internal Medicine*.
17. B. Garibaldi, J. Fiksel, J. Muschelli, M. Robinson, **Masoud Rouhizadeh**, J. Perin, G. Schumock, P. Nagy, J. Gray, H. Malapati, M. Ghobadi-Krueger, T. Niessen, B. Soo Kim, P. Hill, M.S. Ahmed, E. Dobkin, R. Blanding, J. Abele, B. Woods, K. Harkness, A. Gupta. 2021. "Patient Trajectories Among Persons Hospitalized for COVID-19: A Cohort Study." *Annals of Internal Medicine*.
18. Elham Hatef, Xiaomeng Ma, **Masoud Rouhizadeh**, Gurmehar Singh, Jonathan Weiner, Hadi Kharrazi. 2020. "Assessing the Impact of Social Needs and Social Determinants of Health on Healthcare Utilization: Using Patient and Community Level Data." *Journal of Population Health Management*.
19. Lindsay Dickerson, **Masoud Rouhizadeh**, Allan Massie, Mara McAdams-Demarco, Mary Grace Bowring, Dorry Segev, Alicia Cannon, Anthony Guerrerio, Po-Hung Chen, Benjamin Philosophie, Douglas Mogul. 2019. "Language Impairment in Adults with End-stage Liver Disease: a Novel Application of Natural Language Processing." *Nature Digital Medicine*.
20. Elham Hatef, **Masoud Rouhizadeh**, Iddrisu Tia, Elyse Lasser, Felicia Hill-Briggs Jill Marsteller, Hadi Kharrazi. 2019. "Assessing the Availability of Social and Behavioral Determinants Data in Structured and Unstructured Electronic Health Records: A Retrospective Analysis of a Multi-Level Healthcare System." *JMIR Medical Informatics*.
21. Niloofar Mansoori, Mehrnoush Shamsfard, **Masoud Rouhizadeh**. 2012. "Compound verbs in Persian WordNet." *International Journal of Lexicography*, Vol. 25 No. 1, pp. 50-67.

## Book Chapters

1. **Masoud Rouhizadeh**, Margit Bowler, Richard Sproat, and Bob Coyne. 2011. "Data Collection and Normalization for Building the Scenario-Based Lexical Knowledge Resource of a Text-to-Scene Conversion System". In Andreas König, et. al. *Semantic Media Adaptation and Personalization*, pp 378–387, Springer, Berlin, Heidelberg.
2. **Masoud Rouhizadeh**, Bob Coyne, and Richard Sproat. 2011. "Collecting Semantic Information for Locations in the Scenario-Based Lexical Knowledge Resource of a Text-to-Scene Conversion System". In A. König, et. al. (Eds.): KES 2011, Part IV, Lecture Notes in Artificial Intelligence 6884, pp. 378–387. Springer-Verlag Berlin, Heidelberg.
3. Mahsa Yarmohammadi, Mehrnoush Shamsfard, Mahshid Yarmohammadi, **Masoud Rouhizadeh**. 2009. "SBUQA Question Answering System". In Hamid Sarbazi-Azad et. al., *Advances in Computer Science and Engineering*. Communications in Computer and Information Science, vol 6, pp 316–323, Springer, Berlin, Heidelberg.

## Conference Proceedings

All listed publications have undergone rigorous peer review. The conference proceedings are published in top-tier AI and computer science venues, with highly selective acceptance rates of 25% or lower. These venues are recognized for their impact and influence, as evidenced by high h5-index scores based on [Google Scholar Metrics](#) [\[link\]](#).

1. Hossein Rouhizadeh, Mehrnoush Shamsfard, **Masoud Rouhizadeh**. 2021. “Persian SemCor: A Bag of Word Sense Annotated Corpus for the Persian Language”. *Global WordNet Conference (GWC)*, Pretoria, South Africa.
2. Hossein Rouhizadeh, Mehrnoush Shamsfard, **Masoud Rouhizadeh**. 2020. “Knowledge Based Word Sense Disambiguation with Distributional Semantic Expansion for the Persian Language”. *International eConference on Computer and Knowledge Engineering (ICCKE)*.
3. Andriy Mulyar, Elliot Schumacher, **Masoud Rouhizadeh**, Mark Dredze. 2019. “Phenotyping of Clinical Notes with Improved Document Classification Models Using Contextualized Neural Language Models”. *33rd Conference on Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada.
4. Hossein Rouhizadeh, Mehrnoush Shamsfard, **Masoud Rouhizadeh**. 2019. “Knowledge-Based Word Sense Disambiguation with Distributional Semantic Expansion”. *Workshop on Widening NLP in ACL*, Florence, Italy.
5. **Masoud Rouhizadeh**, Kokil Jaidka, Laura Smith, H. Andrew Schwartz, Anneke Buffone, Lyle H. Ungar. 2018. “Identifying Locus of Control in Social Media Language”. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Brussels, Belgium.
6. **Masoud Rouhizadeh**, Arjun Magge, Ari Klein, Abeed Sarker, and Graciela Gonzalez. 2018. “A Rule-based Approach to Determining Pregnancy Time-frame from Contextual Social Media Postings”. *ACM International Digital Health Conference (DH)*, Lyon, France.
7. Kokil Jaidka, Anneke Buffone, Johannes Eichstaedt, **Masoud Rouhizadeh**, and Lyle Ungar. 2018. “Modeling and Visualizing Locus of Control with Facebook Language”. *AAAI Conference on Web and Social Media (ICWSM)*, Stanford, CA, USA.
8. H. Andrew Schwartz, **Masoud Rouhizadeh**, Michael Bishop, Philip Tetlock, Barbara Mellers, and Lyle Ungar. 2017. “Assessing Objective Recommendation Quality through Political Forecasting”. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Copenhagen, Denmark.
9. Ari Z. Klein, Abeed Sarker, **Masoud Rouhizadeh**, Karen O’Connor, Graciela Gonzalez. 2017. “Detecting Personal Medication Intake in Twitter: An Annotated Corpus and Baseline Classification System”. *Conference on Biomedical Natural Language Processing (BioNLP)*, Vancouver, BC.
10. **Masoud Rouhizadeh**, Lyle Ungar, Anneke Buffone, and H. Andrew Schwartz. 2016. “Using Syntactic and Semantic Context to Explore Psychodemographic Differences in Self-reference”. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Austin, TX.
11. **Masoud Rouhizadeh**, Emily Prud’hommeaux, Jan van Santen, and Richard Sproat. 2015. “Measuring idiosyncratic interests in children with autism spectrum disorder”. *Annual Meeting of the Association for Computational Linguistics (ACL)*, Beijing, China.
12. **Masoud Rouhizadeh**, Richard Sproat, and Jan van Santen. 2015. “Similarity Measures for Quantifying Restrictive and Repetitive Behavior in Conversations of Autistic Children”. *Workshop on Computational Linguistics and Clinical Psychology (CLPsych)*, Denver, CO.
13. Emily Prud’hommeaux, Eric Morley, **Masoud Rouhizadeh**, Laura Silverman, Jan van Santen, Brian Roark, Richard Sproat, Sarah Kauper, and Rachel DeLaHunta. 2014. “Computational analysis of trajectories of linguistic development in autism”. *IEEE Spoken Language Technology Workshop (SLT)*, South Lake Tahoe, NV.

14. **Masoud Rouhizadeh**, Emily Prud'hommeaux, Jan van Santen, and Richard Sproat. 2014. "Detecting linguistic restricted interests in autism using distributional semantic models". *Workshop on Computational Linguistics and Clinical Psychology (CLPsych)*, Baltimore, MD.
15. **Masoud Rouhizadeh**, Emily Prud'hommeaux, Brian Roark, and Jan van Santen. 2013. "Distributional semantic models for the evaluation of disordered speech". *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, Atlanta, GA.
16. Bob Coyne, Alex Klapheke, **Masoud Rouhizadeh**, Richard Sproat, and Daniel Bauer. 2012. "Annotation Tools and Knowledge Representation for a Text-To-Scene System". *International Conference on Computational Linguistics (COLING)*, Mumbai, India.
17. Emily Prud'Hommeaux and **Masoud Rouhizadeh**. 2012. "Automatic detection of pragmatic deficits in children with autism". *Workshop on Child, Computer, and Interaction (WOCCI)* at InterSpeech, Portland, OR.
18. **Masoud Rouhizadeh**, Daniel Bauer, Bob Coyne, Owen Rambow, and Richard Sproat. 2011. "Collecting Spatial Information for Locations in a Text-to-Scene Conversion System". *Workshop on Computational Models of Spatial Language Interpretation and Generation (CoSLI)*, Boston, MA.
19. **Masoud Rouhizadeh**, Margit Bowler, Richard Sproat, and Bob Coyne. 2011. "Collecting Semantic Data from Amazon Mechanical Turk for a Lexical Knowledge Resource in a Text to Picture Generating System". *International Conference on Computational Semantics (IWCS)*, Oxford, UK.
20. **Masoud Rouhizadeh**, Mahsa Yarmohammadi, and Mehrnoush Shamsfard. 2010. "Developing the Persian WordNet of Verbs; Issues of Compound Verbs and the Editor". *Global WordNet Conference (GWC)*, Mumbai, India.
21. Mahsa Yarmohammadi, Mehrnoush Shamsfard, and **Masoud Rouhizadeh**. 2008. "Using WordNet in Extracting the Final Answer from Retrieved Documents in a Question Answering System". *Global WordNet Conference (GWC)*, Szeged, Hungary.
22. **Masoud Rouhizadeh**, Mehrnoush Shamsfard, and Mahsa Yarmohammadi. 2008. "Building a WordNet for Persian Verbs". *Global WordNet Conference (GWC)*, Szeged, Hungary.
23. **Masoud Rouhizadeh**, Mostafa Assi, and Mahsa Yarmohammadi. 2007. "Designing Persian Verbs WordNet". *Iranian Conference on Linguistics*, Tehran, Iran.

## Peer-reviewed Abstracts

1. Ruba Sajdeya, Chen Bai, Sebastian Jugl, Ronald L. Ison, Mamoun T. Mardini, Patrick J. Tighe, Kimia Zandbiglari, Hanzhi Gao, Almut G. Winterstein, Thomas A. Pearson, Robert L. Cook, Masoud Rouhizadeh. 2023. "Preoperative cannabis use status ascertainment using natural language processing methods and coded data." *American Medical Informatics Association Annual Symposium (AMIA)*, New Orleans, Louisiana, USA. **Best Poster Award Nominee**
2. Kimia Zandbiglari, Elise Omaki, Farzana Islam Adiba, Wendy C. Shields, Andrea Gielen, Masoud Rouhizadeh. 2023. "Deep Learning and Natural Language Processing Methods for Characterizing Pediatric Falls Events." *American Medical Informatics Association Annual Symposium (AMIA)*, New Orleans, Louisiana, USA.
3. Ruba Sajdeya, Chen Bai, Sebastian Jugl, Ronald L. Ison, Mamoun T. Mardini, Patrick J. Tighe, Kimia Zandbiglari, Hanzhi Gao, Almut G. Winterstein, Thomas A. Pearson, Robert L. Cook, Masoud Rouhizadeh. 2023. "Development and validation of a natural language processing algorithm for extracting preoperative cannabis use status information from unstructured narrative clinical notes." *American Medical Informatics Association Annual Symposium (AMIA)*, New Orleans, Louisiana, USA.

4. Ruba Sajdeya, Chen Bai, Sebastian Jugl, Ronald L. Ison, Mamoun T. Mardini, Patrick J. Tighe Kimia Zandbiglari, Hanzhi Gao, Almut G. Winterstein, Thomas A. Pearson, Robert L. Cook, **Masoud Rouhizadeh**. 2023. "Development and validation of a natural language processing algorithm for extracting preoperative cannabis use status information from unstructured narrative clinical notes." 2023 Consortium for Medical Marijuana Clinical Outcomes Research. Orlando, FL.
5. Ruba Sajdeya, **Masoud Rouhizadeh**, Chen Bai, Sebastian Jugl, Robert L. Cook, Ronald L. Ison, Mamoun T. Mardini, Almut G. Winterstein, Catherine C. Price, Thomas A. Pearson, Patrick J. Tighe. 2023. "Cannabis use and inhalational anesthesia maintenance: A propensity score matched retrospective cohort study." Anesthesiology, San Francisco, CA.
6. Kai-Wen K. Yang, Michelle H. Nguyen, Angie Jelin, **Masoud Rouhizadeh**, Nara Sobreira, Casey Overby Taylor. 2023. "Detecting Phenotypes Among Patients Suspected of Rare Mendelian Disorders." American Medical Informatics Association Informatics Summit, Seattle, WA, USA.
7. Huang W, Ahmed M, Smith S, Hasan MM, **Masoud Rouhizadeh**, Bian J, Kimmel S, Morris EJ, Yang L, Guo J. 2023. "Trajectories of Sacubitril/Valsartan Adherence Among Medicare Beneficiaries with Heart Failure." ISPOR, Boston, MA. Poster.
8. Elise Omaki, Wendy Shields, **Masoud Rouhizadeh**, Pamela Delgado-Barroso, Ruth Stefanos (2022). "Using Natural Language Processing to Understand Contributing Factors of Pediatric Falls". 2022 Annual Meeting of American Public Health Association, Boston, MA, USA.
9. Lemas DJ, **Masoud Rouhizadeh**, Braeden L, Frank S, Wright L, Magalhães M, Xu K, Du X, Parker L, Harle C, Louis-Jaques A, Zhang B, Thompson L, Hogan WR, Modave F. (2022). "Classifying infant feeding status from clinical notes using natural language processing and machine learning". *American Medical Informatics Association Annual Symposium (AMIA)*, Washington, DC, USA.
10. Gray G, Rouhizadeh M, Ahumada L, Richards T, Zirikly A, Hatéf E. (2022). "Application of Natural Language Processing to Identify Social Needs from The Electronic Health Record's Free-Text Notes". *American Medical Informatics Association Annual Symposium (AMIA)*, Washington, DC, USA.
11. Elham Hatéf, **Masoud Rouhizadeh**, Claudia Nau, Fagen Xie, Ariadna Padilla, Lindsay Joe Lyons, Christopher Rouillard, Mahmoud Abu-Nasser, and Douglas Roblin. 2021. "Assessing the Documentation of Social Needs in Electronic Health Records' Unstructured Data: A Collaboration of Johns Hopkins Health System and Kaiser Permanente." *AcademyHealth. Best abstract in Digital Technologies, Data and Analytics.*
12. Elham Hatéf, **Masoud Rouhizadeh**, Claudia Nau, Fagen Xie, Ariadna Padilla, Lindsay Joe Lyons, Christopher Rouillard, Mahmoud Abu-Nasser, and Douglas Roblin. 2021. "A Pilot Study to Improve the Use of Electronic Health Records for Identification of Patients with Social Determinants of Health Challenges: A Collaboration of Johns Hopkins Health System and Kaiser Permanente." *American Medical Informatics Association Annual Symposium (AMIA)*, San Diego, California, USA.
13. Eva Tseng, Jessica Schwartz, **Masoud Rouhizadeh**, Nisa Maruthur. 2020. "Analysis of Primary Care Provider (PCP) EHR Notes for Discussions of Prediabetes Using Natural Language Processing (NLP) Methods". *Journal of Diabetes*, no. 69, supplement 1.
14. Casey Overby Taylor, **Masoud Rouhizadeh**, Franccois Schiettecatte, Howard Levy, Alex Baras, Ada Hamosh, Nara Sobreira. 2019. "Detecting Phenotype Descriptors in Clinical Notes with ClinPhen: An Assessment of Sensitivity in a PhenoDB Patient Cohort". *Annual Meeting of American Society of Human Genetics (ASHG)*, Houston, TX.
15. Andriy Mulyar, Elliot Schumacher, **Masoud Rouhizadeh**, Chris Chute, Mark Dredze. 2019. "Experiments with Pre-Trained Deep Neural Language Models for Clinical NLP: Concept Linking and Semantic Similarity". *National NLP Clinical Challenges (N2C2) Workshop*, Washington, DC.



16. Lindsay Dickerson, **Masoud Rouhizadeh**, Mary Bowring, Douglas Mogul. 2019. "Use of Natural Language Processing (NLP) to Identify Neurocognitive Deficits in End-Stage Liver Disease". *American Journal of Transplantation* Volume 19, Pages 29-30. **Poster of Distinction**
17. Lindsay Dickerson, **Masoud Rouhizadeh**, Allan Massie, Mara McAdams-Demarco, Mary Grace Bowring, Dorry Segev, Alicia Cannon, Anthony Guerrero, Po-Hung Chen, Benjamin Philosophie, Douglas Mogul. 2019. "Language Impairment in Adults with End-Stage Liver Disease: a Novel Application of Tools from Natural Language Processing". *Gastroenterology*. Volume 156, Issue 6, Supplement 1, Page S-1369.
18. **Masoud Rouhizadeh**, Elham Hatf, Mark Dredze, Christopher Chute, Hadi Kharrazi. 2018. "Identifying Social Determinants of Health from Clinical Notes: a Rule-Based Approach". AMIA NLP working group pre-symposium. San Francisco, CA.
19. Michelle S. Horner, **Masoud Rouhizadeh**. 2018. "Homicidal and suicidal ideation as a chief complaint: considering the influence of news events using a large, de-identified electronic medical record dataset." Society for Prevention Research, Washington, DC.
20. **Masoud Rouhizadeh**, Arjun Magge, Ari Klein, Abeed Sarker, Graciela Gonzalez. 2017. "Detecting Gestation Period Using Social Media Data Analysis". Penn Biomedical Postdoctoral Research Symposium, Philadelphia, PA.
21. **Masoud Rouhizadeh**, H Andrew Schwartz. 2016. "Age and Gender Differences in Self-reference in Verb Categories." Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL), Philadelphia, PA.
22. **Masoud Rouhizadeh**, Richard Sproat, and Jan van Santen. 2015. "Computational Semantic Analysis of Restrictive and Repetitive Behavior in Language Samples of Children with Autism." *International Meeting for Autism Research (IMFAR)*, Salt Lake City, UT.
23. **Masoud Rouhizadeh**, Jan van Santen, Richard Sproat, Kyle Gorman, Peter Heeman, Alison Presmanes Hill, Steven Bedrick, Emily Tucker Prud'hommeaux, Géza Kiss. 2014. "Discourse Marker Use in ASD and Typical Development." *Pacific Northwest NLP Workshop: NW-NLP*, Redmond, WA.
24. **Masoud Rouhizadeh**, Jan van Santen, Richard Sproat, Peter Heeman, Alison Presmanes Hill, Steven Bedrick, and Emily Prud'hommeaux. 2014. "Children's Differing Patterns of Discourse Marker Use in ASD and Typical Development." *International Meeting for Autism Research (IMFAR)*, Atlanta, GA.
25. Emily Prud'hommeaux, **Masoud Rouhizadeh**, Brian Roark, and Jan van Santen. 2013. "Identifying Unexpected and Inappropriate Words in ASD Language Samples." *International Meeting for Autism Research (IMFAR)*, San Sebastian, Spain.
26. **Masoud Rouhizadeh**, Richard Sproat and Bob Coyne. 2012. "Collecting Spatial Information for Locations in a Text-to-Scene Conversion System". *Pacific Northwest NLP Workshop: NW-NLP*, Redmond, WA.

## Preprint

1. Kimia Zandbiglari, Hamid Reza Hasanzadeh Hasanzadeh, Pareeta Kotecha Kotecha, Ruba Sajdeya, Amie J. Goodin, Tianze Jiao, Farzana Adiba, Mamoun T. Mardini, Jiang Bian, and **Masoud Rouhizadeh**. 2023. "A Natural Language Processing Algorithm for Classifying Suicidal Behaviors in Alzheimer's Disease and Related Dementia Patients: Development and Validation Using Electronic Health Records Data." *medRxiv*.
2. S. Liu, A. Wen, L. Wang, H. He, S. Fu, R. Miller, A. Williams, D. Harris, R. Kavuluru, M. Liu, N. Abu-el-Rub, D. Schutte, R. Zhang, **Masoud Rouhizadeh**, J. D. Osborne, Y. He, U. Topaloglu, S. S. Hong, J. H. Saltz, T. Schaffter, E. Pfaff, C. G. Chute, T. Duong, M. A. Haendel, R. Fuentes, P. Szolovits, H. Xu, H. Liu. 2021. "An Open Natural Language Processing Development

Framework for EHR-based Clinical Research: A case demonstration using the National COVID Cohort Collaborative (N3C)". *arXiv*.

3. Jingqi Wang, Huy Anh Pham, Frank Manion, **Masoud Rouhizadeh**, Yaoyun Zhang. 2020. "COVID-19 SignSym – A fast adaptation of general clinical NLP tools to identify and normalize COVID-19 signs and symptoms". *arXiv*.
4. Rachel Dorn, Alicia L Nobles, **Masoud Rouhizadeh**, Mark Dredze. 2020. "Examining the Feasibility of Off-the-Shelf Algorithms for Masking Directly Identifiable Information in Social Media Data". *arXiv*.
5. B. Garibaldi, J. Fiksel, J. Muschelli, M. Robinson, **Masoud Rouhizadeh**, P. Nagy, H. Malapati, M. Ghobadi-Krueger, T. Niessen, B. Soo Kim, P. Hill, M.S. Ahmed, E. Dobkin, R. Blanding, J. Abele, B. Woods, K. Harkness, D. Thiemann, M.G. Bowring, A. Shah, M.C. Wang, K. Bandeen-Roche, A. Rosen, S. Zeger, A. Gupta. 2020. "Patient trajectories and risk factors for severe outcomes among persons hospitalized for COVID-19 in the Maryland/DC region". *medRxiv*.
6. Joao Sedoc, Derry Wijaya, **Masoud Rouhizadeh**, Andrew Schwartz, Lyle Ungar. "Deriving Verb Predicates By Clustering Verbs with Arguments". *arXiv*.

RESEARCH  
SUPPORT

**Current**

P30AG066506      NIH-NIA      Smith (PI)      06/2021-04/2025  
*1Florida Alzheimer's Disease Research Center*  
Role: Segment PI

P30 grant that supports multiple investigators to enhance multidisciplinary approaches and collaborative research efforts focused on a common research problem or goal. This project focuses on research in the area of Alzheimer's.

U18DP006711      CDC      Shao (PI)      09/2022-08/2027  
*Assessing Barriers and Facilitators for Participating Structured Lifestyle Intervention and Its Real-world Effectiveness and Cost-effectiveness among US Veterans*  
Role: Co-I

This project aims to identify barriers and facilitators to the enrollment and completion of the National Diabetes Prevention Program, or NDPP, and evaluate its long-term effectiveness and cost-effectiveness. The research team will leverage both structured and unstructured data from the Veterans Administration to identify individual and system-level factors associated with the enrollment and completion of the MOVE! program. They will specifically focus on social determinants of health that may be associated with racial/ethnic inequity in program enrollment and completion.

1Florida Alzheimer's Disease Research Center (ADRC)      Rouhizadeh (PI)      01/2022-11/2024  
*Identifying suicidal behaviors using Natural Language Processing in patients with Alzheimer's disease and related dementias*  
Role: PI

The overarching goal of this study is to apply, evaluate, and refine a natural language processing-based approach to identify suicide attempts and suicidal ideation in patients with Alzheimer's disease and related dementias (AD/ADRD). We also plan to determine the epidemiology of suicide attempts and suicidal ideation using the expanded classification and evaluate differential detection of suicide attempts and suicidal ideation via codified vs. NLP-based approaches.

1Florida Alzheimer's Disease Research Center (ADRC)      McDonough (PI)      01/2022-11/2024  
*Identification and Characterization of Blood Pressure Control and Racial Impacts on Alzheimer's Disease and Related Dementias Risk*  
Role: Co-I

Hypertension is one of the potentially modifiable risk factors for Alzheimer's disease and related dementias. African Americans and Hispanics have higher rates of Alzheimer's disease and related dementias, and lower rates of blood pressure control compared to non-Hispanic whites. This project will validate Alzheimer's disease and related dementias models in African Americans and Hispanics and provide an improved understanding of the role of blood pressure

control in risk for Alzheimer’s disease and related dementias, allowing high risk patients to be identified sooner, and targeted to precision treatment regimens.

UF Health Puri (PI) 11/2023-10/2025

Can knowledge of CYP2C19 genotyping help determine patients with GERD that are better managed surgically

Role: Co-I

This retrospective review proposes to: 1) Assess the prevalence of genetic polymorphisms in patients with GERD in the Northern Florida population; 2) Determine the degree of polymorphisms in medical and surgical patients with GERD; 3) Develop a predictive model encompassing anatomic, physiologic, and genetic factors that may be applied clinically to better manage GERD patients. This will be the first study to compare CYP2C19 polymorphisms in adult surgical versus medical patients with GERD.

## Pending

PRO00054645 NIH NIDA Rouhizadeh (PI) 07/2024-06/2029

*Utilizing Natural Language Processing in Electronic Medical Records to Investigate the Relationship between Cannabis Use and Health Outcomes*

Role: PI

Cannabis use has increased in prevalence across most of the United States since recreational or medicinal use was legalized in all but eight states by 2023. This proposal investigates the impacts of cannabis use on mental health outcomes among young adults, examining its relationships with new-onset psychosis, self-harm, and substance use disorders, while also assessing how social determinants like income and housing modify these associations. Findings will provide real-world evidence to inform clinical practice, population health studies, and policy regarding rising cannabis use and associated mental health risks in this demographic.

PRO00051095 NIH NCATS Liu (PI) 04/2024-03/2029

Open Health Natural Language Processing Collaborative towards Fair and Inclusive Clinical and Translational Research

Role: Co-I

The project aims to establish a privacy-preserving Open Health Natural Language Processing (OHNLP) ecosystem for sharing, testing, integrating, and deploying NLP solutions to identify special or underrepresented populations. By enhancing the enterprise data warehouse for research with text analytics capabilities through the open text analytics ecosystem, we demonstrate the utility of OHNLP in supporting fair and inclusive clinical and translational research in rural health, disability, and rare diseases.

NIH NHLBI McDonough (PI) 04/2024-03/2029

*Leveraging Electronic Health Records to Improve Identification and Risk Prediction of Treatment Resistant*

Role: Co-I

Our research goal is to create algorithms and methodologies within the electronic health record (EHR) to identify and predict patients at high risk for cardiovascular disease. Our overall objective for this application is to build on the results from Dr. McDonough’s (PI) NHLBI K01 award, by developing improved algorithms that identify apparent treatment resistant hypertension (aTRH) patients at increased risk for cardiovascular disease and facilitate their future implementation in EHRs with disparate data models.

DoD Mardini (PI) 03/2024-09/2024

*Leveraging Large Language Models for Summarizing Electronic Clinical Note*

Role: Co-I

Healthcare professionals face overwhelming data volumes in electronic records, hampering extraction of essential details. We propose automatically creating summaries of complex clinical narratives via Large Language Models, improving information retrieval and decisions. Summaries will be iteratively refined through physician-in-the-loop evaluation assessing conciseness, relevance, readability, and incorporating expert feedback. This research can offer insights on using language models for clinical summarization, develop physician-validated standards for implementation,

and inform best practices for integrating these tools into healthcare systems.

DoD - Johns Hopkins University                      Rouhizadeh (PI)                      09/2024-09/2026  
*Assessing the Impact of Symptom Burden and Treatment on Disability and Progression in Multiple Sclerosis*  
Role: PI

We will develop large language models to identify hidden symptoms of multiple sclerosis from free text clinical notes within electronic health records. We will build and evaluate large language models on Johns Hopkins MS Center visit notes to expand data collection of fatigue, depression, cognitive dysfunction, pain, anxiety, spasticity, and bowel/bladder dysfunction. This will allow us to conduct a case-control study for assessing the impact of symptom burden and treatment on disability and progression in MS.

## Completed

1R01DA050676                      NIH NIDA                      Lo-Ciganic (PI)                      07/2021-00/2023  
*Developing and Evaluating a Machine-Learning Opioid Prediction & Risk-Stratification E-Platform (DEMONSTRATE)*  
Role: Co-I

The proposed study aimed to harness advanced natural language processing and longitudinal neural network approaches to build on our previously developed machine-learning prediction algorithms to identify patients at risk for opioid overdose or opioid use disorder. We developed the prediction tool using all-payer electronic health records (EHR), Medicaid claims, and Medicaid claims linked with EHR data from the One Florida Clinical Research Consortium and translate the risk prediction algorithms into a clinical decision support platform integrated into the EHR system to identify patients at high risk of overdose and opioid use disorder.

1R01MD015844                      NIH NIMHD                      Weiner, Hatef (PIs)                      07/2021 - 03/2023  
*Development, Piloting and Dissemination of an Integrated Clinical and Social Multi-level Decision Support Platform to Address Social Determinants of Health Among Minority Populations in Baltimore City*  
Role: Co-I

With the intent of improving care for minority and disadvantaged populations with chronic diseases, this project developed methods and tools to better integrate available digital information regarding social determinants of health (SDOH) into providers' electronic health records (EHRs). The study developed, piloted, evaluated, and widely disseminated a multi-level EHR-integrated clinical decision support system to help identify, manage, and refer patients with high chronic disease burden who also had high levels of modifiable SDOH challenges.

5U01FD005942-05                      FDA CERSIs                      Kharrazi, Weiner (PIs)                      09/2020 - 09/2021  
*Assessing Disparities in Occurrence and Outcomes of Type 2 Diabetes ADEs in Minority Populations using Real World Administrative Claims and Electronic Health Records*  
Role: Co-I

This project aimed to improve the identification of severe hypoglycemia (SHG) in ambulatory EHRs and claims. We compared SHG events across races/ethnicities and various SDH factors, and then identified key disparity factors associated with an increased likelihood of SHG among African American patients. We also discovered contextual patterns associated with higher rates of SHG among different minority and special-need populations.

HSRP20202805                      PCORI                      Turchin, Ford (PIs)                      11/2019 - 09/2021  
*Using Topic Segmentation to Improve Concept Parsing and Identification of Negation in Extraction of Data from EHRs*  
Role: Project Leader

A major challenge when using EMR data for research was that a large fraction of the data was contained in free-text documents. Even with modern NLP methods, it could be difficult to extract information when the topic was switched mid-sentence. The goal of this project was to build NLP for topic segmentation to help with better clinical concept extraction.

1R01MH124724      NIH NIMH      Kharrazi (PI)      09/2020 - 09/2021  
*Advancing Maryland's Statewide Suicide Data Warehouse to Improve Individual and Population-level Mortality Prediction and Prevention*

Role: Co-I

The aim of developing Maryland Suicide Data Warehouse (MSDW) was to link a diverse set of data sources and use various layers of risk factors to predict suicide death on a generalizable population. MSDW included data on 5+ million Maryland residents spanning from 2012 to 2017. MSDW contained hospital and emergency discharges, health information exchange (HIE) data, commercial insurance claims, medical examiner data, electronic health record (EHR) data from 5 select health systems, and place-based social determinants of health data covering various domains such as housing, employment, education, income and crime. This study leveraged the unique data types linked by MSDW to improve suicide death predictions among generalizable patient populations.

1R56MH117560      NIH NIMH      Kharrazi, Wilcox (PIs)      09/2018 - 09/2021  
*Addressing Suicide Research Gaps: Understanding Mortality Outcomes in the Mid-Atlantic Region*

Role: Co-I

This study conducted data linkage and informatics approaches to utilize existing resources to improve suicide risk identification and prevention. The project also offered essential measures for providers and payers to reduce suicide events in their systems. The methods were assessed against and shared as a resource with other states/regions to establish a similar analytics framework to advance suicide prevention.

UF Informatics Institute (UFII)      Rouhizadeh, Brown (MPI)      01/2022-01/2022  
*Diagnosing suicidal behaviors in postpartum mothers using natural language processing*

Role: MPI  
In this project we refined an existing NLP text mining and heuristic rule-based algorithm that identified suicide attempts and behaviors in postpartum mothers and to apply the refined NLP tool and evaluate demographic and clinical characteristics associated with a suicide attempt and suicide-related behaviors.

UF Informatics Institute (UFII)      McDonough, Cavallari (MPI)      01/2022-01/2022  
*Utilizing Artificial Intelligence to Identify Bleeding Risk Predictors with Newer P2Y12 Receptor Inhibitors*

Role: Co-I

This project aimed to identify patient-specific factors predictive of bleeding risk with prasugrel or ticagrelor after PCI through AI approaches; and to derive a novel score for predicting bleeding risk with prasugrel and ticagrelor that could be integrated into the EHR to provide point-of-care risk assessment.

Johns Hopkins ICTR Nexus Award      Kharrazi (PI)      06/2020-05/2021  
*Behavioral, Social and Systems Science: Extracting Social Science Data from Epic Electronic Health Record System*

Role: Co-I

This phase of the project completed the extraction of social and behavioral determinants of health from Epic's EHR as well as integrated place-based social determinants of health into JHMI's research platforms.

JHU Institute for Clinical and Translational Research      Chen (PI)      01/2021-01/2022  
*Socioeconomic Determinants of Alcohol Addiction Treatment for Alcohol Use Disorder in Patients with Alcoholic Hepatitis*

Role: Co-I

The burden of alcohol-related liver disease continued to increase with rising alcohol consumption, disproportionately affecting disadvantaged groups. We explored the socioeconomic factors that were associated with under-utilization of alcohol addiction treatment and delineated the areas of health equity that could be barriers to it.

JHU Precision Medicine Analytics Platform      Rouhizadeh (PD)      04/2020-03/2021  
*AI and NLP methods from extracting COVID-19 signs and symptoms from free text clinical notes*

Role: Project Lead

Statistical modeling of patient trajectories among persons hospitalized for COVID-19 was critical to understand the disease progression and provide better clinical care. The JHU clinical and biostatistics experts created a model for

COVID-19 progression to severe disease or death. Factors included demographics, vitals, comorbid conditions as well as presenting symptoms extracted from hospital notes using NLP methods.

JHU Precision Medicine Analytics Platform                      Rouhizadeh (PD)                      03/2019-02/2021  
*Semantic tagging and concept extraction from free-text clinical notes in the JHMI Health Systems*  
Role: Project Lead

JHMI EHR system already contained 350m clinical notes and on average 70k notes were added daily. Our enterprise-level big-data concept extraction tool was designed to extract clinical concepts and convert imperceptible unstructured data into a structured format, linked to the NIH UMLS standardized vocabulary.

JHU Institute for Clinical and Translational Research                      Cai (PI)                      01/2021-01/2022  
*Association of Race, Ethnicity, and Income with Risk of Vision Loss from Diabetic Macular Edema: Evaluating the Role of Lapses in Care*  
Role: Co-I

This project examined the association of race, ethnicity, and income with risk of vision loss from diabetic macular edema using NLP approaches. Identifying lapses in care as a modifiable risk factor underlying health inequities in diabetic macular edema helped prioritize clinical interventions and promote health equity.

Johns Hopkins – Kaiser Permanente Collaboration Award                      Hatf, Nau (PIs)                      03/2020-06/2021  
*Applying Unstructured EHR Notes and Structured Clinical and Community Data to Identify Patient Social Needs and Determinants*  
Role: Co-I

The overarching goal of this project was to develop and apply new practical tools and approaches for measuring Social Determinants of Health using unstructured (e.g., clinician notes) and structured EHR data and information derived from neighborhood databases.

JHU Precision Medicine Analytics Platform                      Rouhizadeh (PD)                      08/2019-07/2020  
*Standardizing clinical notes within into HL7 Clinical Document Architecture (CDA)*  
Role: NLP Lead

The goal of this project was to build a parser to convert Epic-generated Rich Text Format (RTF) notes into semi-structured HL7 Clinical Document Architecture (CDA), and preserve valuable information stored in RTFs such as section headers, section boundaries, and tabular formats. We applied this process to the enterprise-level repository of all 400m clinic notes in JHMI.

JHU Institute for Data Intensive Engineering and Science                      Weiner (PI)                      04/2018-03/2019  
*Harnessing Big Data for Population Health: Advancing NLP Techniques to Extract Social/Behavioral Risk Factors from Free Text within Large Electronic Health Record Systems*  
Role: Co-I

We built on the machine learning tools developed using a mix of pattern matching techniques for note categorization and deep learning methods. Our goal in this pilot was to establish the viability of these methods for social determinants of health (SDH) data and to extract a first-pass corpus of SDH factors.

Johns Hopkins Core Facilities                      Rouhizadeh (PD)                      01/2018-12/2019  
*Center for Clinical Data Analysis (CCDA) Core Coins*  
Role: NLP Lead

Our goal in this program was to provide junior researchers and unfunded study teams the opportunity to utilize NLP tools and methodologies to analyze clinical notes to support their research.

Johns Hopkins ICTR Nexus Award                      Kharrazi (PI)                      12/2017-12/2018  
*Behavioral, Social and Systems Science: Extracting Social Science Data from Epic Electronic Health Record System*  
Role: Co-I

In this phase of the project, we documented the social determinant measure available, provided a guide to researchers wishing to integrate them into studies, and made recommendations on how to integrate new measures going forward.

TRT0048      Templeton Religion Trust      Seligman (PI)      11/2015-02/2017  
*Measuring well-being using big data, social media, and language analyses*  
Role: Postdoctoral Research Associate

Founding the World Well-Being Project at the University of Pennsylvania, we developed new techniques for measuring psychological and medical well-being based on language in social media. We investigated how psychosocial processes affected health and happiness, and developed unobtrusive well-being measures to supplement expensive survey methods.

R01DC012033      NIH NIDCD      van Santen (PI)      09/2012-11/2015  
*Computational Characterization of Language Use in ASD*  
Role: Graduate Research Assistant

In this research, we developed and validated new NLP methods that automatically measured language characteristics of autism, based on raw transcripts of natural language samples. The objective was to improve the analysis of language samples by enhancing efficiency, reliability, and richness of information extracted.

RI-0904361      NSF      Hirschberg (PI)      09/2009-09/2012  
*Collaborative Research: From Text to Pictures*  
Role: Graduate Research Associate

The research teams developed new theoretical models and technology to automatically convert descriptive text into 3D scenes representing the text's meaning. We did this via creating VigNet, a resource they created from existing sources (PropBank, WordNet, FrameNet) and from automated mining of Wikipedia and other unannotated text.

## TEACHING

### University of Florida

*Assistant Professor* - AI in the Health Sciences Initiative

- Introduction to Artificial Intelligence in Pharmacy      Fall 2023
- Applied Data Analysis, Interpretation and Reporting of Findings      Fall 2023
- Introduction to Pharmaceutical Outcomes and Policy      Spring 2022, 2023
- Measurement in Pharmaceutical Outcomes and Policy Research      Fall 2022

### Johns Hopkins University

*Adjunct Assistant Professor* - Biomedical Informatics and Data Science (BIDS)

- Natural Language Processing in the Health Sciences      Spring 2019-2023
- Unstructured Data Mining to Address Novel Infectious Diseases      Summer & Fall 2020
- NLP Methods in the Johns Hopkins Precision Medicine Analytics Platform      Fall 2018, 2019
- Natural Language Processing in Health Care Methodologies      Spring 2020, 2021

*Faculty Instructor* - Center of Excellence Analytics in Medicine Program (CAMP):

- NLP Methods in the Johns Hopkins Precision Medicine Analytics Platform      Fall 2018, 2019

*Instructor* - Leadership in Analytics and Data Science (LEADS):

- Natural Language Processing in Health Care Methodologies      Spring 2020, 2021

### Stony Brook University

*Instructor*:

- Symposium on Natural Language Processing for Social Science      Fall 2016

### University of Tehran

*Lecturer*:

- General English Fall 2008

## University of Applied Science and Technology

*Lecturer*

- Translation of Press Texts Fall 2008
- Contrastive Analysis Fall 2008
- Audio Video Translation Fall 2008
- English Composition Spring 2008
- Translation of Legal Texts Spring 2008
- Introduction to Linguistics Spring 2008
- Principles of Writing Winter 2008
- Translation of Economic Texts Winter 2008
- English Grammar II Winter 2008
- Reading Comprehension Fall 2007
- English Grammar I Fall 2007
- English Conversation Fall 2007

## Azad University of Ramhormoz

*Lecturer:*

- General English II Spring 2007, Summer 2007
- General English I Fall 2006, Winter 2007

CURRICULUM  
DEVELOPMENT

## University of Florida

*College of Pharmacy AI Committee:*

- **Core Faculty:** College of Pharmacy AI Certificate
- **Course Co-Coordinator:** Introduction to Artificial Intelligence in Pharmacy

*Department of Pharmaceutical Outcomes & Policy:*

- **Track Leader:** AI in Pharmaceutical Outcomes and Policy Research PhD/MSc Specialization

## Johns Hopkins University

*Biomedical Informatics and Data Science:*

- **Core Faculty:** JHU Clinical Data Science Initiative
- **Course Co-Director:** Natural Language Processing in the Health Sciences
- **Course Co-Director:** Preparing Multidisciplinary Learners for Amazon Alexa

## Stony Brook University

*Department of Linguistics:*

- **Advisory Board Member:** Computational Linguistics Master's Program



## University of Florida

*AI in the Health Sciences Initiative:*

- Shobhan Kumar (Ph.D): Currently: *Postdoc at UF*
- Farzana Islam Adiba (M.Sc.): Currently: *PhD student at UF*
- Kimia Zandbiglari (M.Sc.): Currently: *PhD student at UF*
- Ruba Sajdeya (MD): Currently: *PhD student at UF*
- Matthew Muschett (PharmD): Currently: *PhD student at UF*
- Cameron Thomas (PharmD): Currently: *PhD student at UF*
- Wenxi Huang (M.Sc.): Currently: *M.Sc. student at UF*

## Johns Hopkins University

*Division of Health Sciences Informatics:*

- Alex Simpson (MD): Currently: *Neuroimmunology Fellow at Johns Hopkins Medicine*
- Helen Ting He (M.Sc.): Currently: *PhD student at Johns Hopkins Medicine*
- Santiago Alvarez Arango (M.D.): Currently: *Clinical Fellow at Johns Hopkins Medicine*
- Kerry Smith (M.Sc.): Currently: *Senior Advisor at Northwell Health*

## Stony Brook University

*Department of Computer Science:*

- Aman Raj (CS M.Sc.): Now: *Senior Software Engineer at Google.*
- Rahool Paliwal (CS M.Sc.): Now: *Senior System Software Engineer at NVIDIA.*
- Danny Bernstein (B.Sc.): Now: *Harvard College.*
- Rishabh Agrawal (CS M.Sc.): Now: *Senior SDE at Amazon AWS AI.*

## Oregon Health & Science University

*Center for Spoken Language Understanding:*

- Margit Bowler (B.A. Reed College): Now: *Annotation Project Manager & Linguist at Apple.*
- Heather Sidener (B.A. Reed College): Now: *Head, Clinical Medicine Unit at OHSU*

## Merit Reviewing for Funding Agencies

- NSF, Small Business Innovation Research - SBIR/STTR - Reviewer 2023-present
- Patient-Centered Outcomes Research Institute (PCORI) - Merit Reviewer 2021-present
- UF Office of Research, Faculty Research Enhancement Program - Reviewer 2023-present
- JHU Institute for Clinical & Translational Research - Technical Reviewer 2018-present

## Area Chair

- NeurIPS - Machine Learning for Health (ML4H) 2021-2023
- Association for Computational Linguistics - Cognitive Modeling & Psycholinguistics 2020

## Co-chair

- Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data (RaPID) 2021
- Diversity & Inclusion Subcommittee - Association for Computational Linguistics (ACL) 2020
- The 4th Pacific Northwest Regional NLP Workshop (NW-NLP) at Amazon, Seattle, WA. 2016

## University Governance

- AI Specialty track leader, UF Pharmaceutical Outcomes & Policy 2022-present
- UF COP AI Committee Member 2021-present
- UF Research Computing Advisory Committee Member. 2011-present

## Organizing Committees

- The 11th Conference on Information and Knowledge Technology (IKT2020) 2020
- Diversity & Inclusion - NAACL 2019
- Conference of the International Speech Communication Association (Interspeech). 2012
- The 49th Annual Meeting of the Association for Computational Linguistics (ACL-HLT) 2011

## Special Interest Group Lead

- Capturing Socio-Behavioral Determinants of Health Data in Johns Hopkins EHR 2021
- IEEE Spoken Language Technology Workshop (SLT) 2014

## Journal Editorial Board

- Nature – Humanities and Social Sciences Communications 2022-present
- Frontiers in Drug Safety and Regulation 2021-present

## Journal Reviewing

- Journal of Healthcare Informatics Research 2022-present
- Nature Humanities and Social Sciences Communications 2022-present
- Nature Partner Journals - Digital Medicine 2022-present
- Journal of Medical Internet Research 2021-present
- Journal of Biomedical Informatics 2021-present
- Journal of Natural Language Engineering 2020-present
- Journal of the American Medical Informatics Association 2020-present
- Patterns - Cell Press 2020-present
- PLOS ONE 2019-2021
- Information Processing and Management (Elsevier IPM) 2018-present
- ACM Transactions on Asian Language Information Processing (TALLIP) 2013-2020
- Language Resources and Evaluation journal (LREV) 2012-2016

## Reviewer and Program Committee

- Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data 2022-2024
- Language Resources and Evaluation (LREC) 2018, 2020, 2024
- American Medical Informatics Association (AMIA) Annual Symposium 2019-2023
- AI Techniques in Interaction-centric Autism Research and Diagnosis 2023
- AcademyHealth Annual Research Meeting (ARM) 2021-2022
- AMIA NLP Working Group 2022-2023
- Association for Computational Linguistics (ACL-IJCNLP) 2020-2021
- North American Association for Computational Linguistics (NAACL) 2016, 2018-2019, 2021

- Annual Meeting of the Cognitive Science Society (CogSci) 2020-2021
- Computational Linguistics and Clinical Psychology (CLPsych) 2017-2021
- Empirical Methods in Natural Language Processing (EMNLP) 2017-2020
- IEEE Signal Processing Society - Acoustics, Speech, and Signal Processing (ICASSP) 2020
- American Medical Informatics Association (AMIA) Informatics Summit 2018-2020
- Negative Results in NLP 2020
- Computational Natural Language Learning (CoNLL) 2019
- Speech and Language Processing for Assistive Technologies (SLPAT) 2019
- Abusive Language Online (ALW) 2017-2019
- Pacific Northwest Regional NLP Workshop (NW-NLP) 2014, 2016, 2018
- IEEE Human-Centered Computational Sensing (HCCS) 2018
- Pacific Symposium on Biocomputing (PSB) 2018
- Social Media Mining for Health Applications (SMM4H) 2017
- Biomedical Natural Language Processing (BioNLP) 2017
- International Conference on Computational Linguistics (COLING) 2016

## Membership

- American Association of College of Pharmacy (AACP) 2021-present
- Open Health Natural Language Processing Consortium (OHNLP) 2020-present
- NIH National COVID Cohort Collaborative (N3C) 2020-present
- Health Level Seven International (HL7) Working Group 2019-present
- Observational Health Data Sciences and Informatics (OHDSI) 2017-present
- American Medical Informatics Association (AMIA) NLP working group 2017-present
- The New York Academy of Sciences 2015-present
- Association for Computing Machinery (ACM) 2013-present
- Association for Computational Linguistics (ACL) 2010-present
- Global WordNet Association (GWA) 2007-present
- International Society for Autism Research (INSAR) 2013-2015
- International Speech Communication Association (ISCA) 2014-2015

## LANGUAGES

- Persian Native
- English Fluent
- Arabic Intermediate
- Old Persian Basic
- Middle Persian Basic
- Avestan Basic
- Italian Elementary

COMPUTER  
SKILLS

**Languages:**

- Python
- R
- SQL
- C
- Java
- Prolog
- XML/CSS
- UNIX shell script.
- 

**NLP, Deep Learning, and Data Science tools:**

- PyTorch
- TensorFlow
- scikit-learn
- spaCy
- cTAKES
- scispaCy
- medspacy
- flairNLP
- NLTK
- Stanford CoreNLP
- ARK TweetNLP
- SyntaxNet
- Word2Vec
- Gensim
- FastText
- Lucene
- Hadoop
- Map/Reduce
- Condor
- MySQL
- Weka
- Praat
- OpenFST
- Lextools
- Thrax
- HTK
- GIZA++
- Moses
- E-Prime
- Jira

PRESENTATIONS

“Developing a Machine-Learning-Based Prediction Model for Diabetes Duration Using Information from Electronic Health Records”. *American Diabetes Association*. May 2023.

“Trajectories of Sacubitril/Valsartan Adherence Among Medicare Beneficiaries with Heart Failure”. *ISPOR*. Boston, MA, Mar 2023.

“Detecting Phenotypes Among Patients Suspected of Rare Mendelian Disorders”. *American Medical Informatics Association Informatics Summit*. Seattle, WA, Feb 2023.

“Investigating AI and Natural Language Processing Methods for Extracting Social Determinants of Health from Unstructured Data in Electronic Medical Records”. *Rita Kobb Nursing and Health Informatics Symposium*. Gainesville, FL, Feb 2023.

“Exploring Deep Learning and AI Techniques for Extracting Information from Unstructured Data in Electronic Medical Records”. *Johns Hopkins All Children’s Hospital*. St. Petersburg, FL, Jan 2023.

“AI and NLP Methods to Extract Socio-Behavioral Data from Unstructured Electronic Medical Records”. *UF AI2HEAL DATATHON*. Gainesville, FL, Nov 2022.

“Using Natural Language Processing to Understand Contributing Factors of Pediatric Falls”. *American Public Health Association*. Boston, MA, Nov 2022.

“Application of Natural Language Processing to Identify Social Needs from The Electronic Health Record’s Free-Text Notes”. *American Medical Informatics Association*. Washington, DC, Nov 2022.

“Classifying infant feeding status from clinical notes using natural language processing and machine learning”. *American Medical Informatics Association*. Washington, DC, Nov 2022.

“Identifying Social Determinants of Health (SDoH) Using Natural Language Processing”. *American Medical Informatics Association*. Invited Speech. Washington, DC, Oct 2022.

“Cannabis use and inhalational anesthesia maintenance: A propensity score matched retrospective cohort study”. *Anesthesiology*. Seminar Presentation. San Francisco, CA.

“Deep Learning and Natural Language Processing Methods for Mining Electronic Medical Records”.

*Loyola University Chicago Health Informatics Seminar Series*. Chicago, IL, Dec 2021.

“AI and natural language processing methods for extracting clinical concepts and identifying social determinants of health from clinical notes”. *The University of Tennessee Health Science Center*. Memphis, TN, Dec 2021.

“Harnessing Big Data for Population Health: Advancing NLP Techniques to Extract Social-Behavioral Risk Factors from Free Text within Large Electronic Health Record Systems”. *OHDSI NLP working group*. Oct 2021.

“Harnessing Big Data for Population Health: Advancing NLP Techniques to Extract Social-Behavioral Risk Factors from Free Text within Large Electronic Health Record Systems”. *N3C NLP working group*. Nov 2021.

“Deep Learning and Natural Language Processing Methods for Mining Electronic Medical Records”. *University of Florida*. Gainesville, FL, June 2021.

“Deep Learning and Natural Language Processing Methods for Mining Electronic Medical Records”. *Mayo Clinic*. Rochester, MN, June 2021.

“Deep Learning and NLP Methods for Mining Electronic Medical Records”. *The University of Texas Health Science Center at Houston*. Houston, TX, May 2021.

“Natural Language Processing in the Health Sciences”. *Columbia University Irving Medical Center*. New York, NY, April 2021.

“Deep Learning and NLP Methods for Mining Electronic Medical Records”. *Columbia University Irving Medical Center*. New York, NY, April 2021.

“Natural Language Processing Methods to Extract Socio-Behavioral Data from Unstructured EHR”. In *Symposium on Capturing Socio-Behavioral Determinants of Health Data in Johns Hopkins EHR*, Johns Hopkins University. Baltimore, MD, March 2021.

“Using Natural Language Processing Methods to Identify Discussions of Prediabetes in EHR notes”. *Johns Hopkins Medicine*. Baltimore, MD, December 2020.

“Deep Learning and NLP Methods for Mining Electronic Medical Records”. *Center for Language and Speech Processing - Johns Hopkins University*. Baltimore, MD, October 2020.

“How will AI influence the engineering profession?”. *Chalmers University of Technology*. Gothenburg, Sweden (remote presentation), December 2019.

“Deep Learning and NLP Methods in Healthcare”. *Chalmers University of Technology*. Gothenburg, Sweden (remote presentation), December 2019.

“NLP and Deep Learning Models for Extracting Housing Challenges from EHR”. *AMIA 2019 Public Health Informatics Working Group*. Washington, DC, November 2019.

“Experiments with Pre-Trained Deep Neural Language Models for Clinical NLP: Concept Linking and Semantic Similarity”. *AMIA-affiliated Workshop on Challenges in NLP for Clinical Data*. Washington, DC, November 2019.

“Deep Learning and NLP Methods for Mining Electronic Medical Records”. *NYU Langone Health*. New York, NY, October 2019.

“Identifying Social Determinants of Health from Clinical Notes: a Rule-Based Approach”. *AMIA NLP working group pre-symposium*. San Francisco, CA, December 2018.

“Applications of Natural Language Processing in Identifying Language Disorders”. *The Johns Hopkins University School of Medicine*. Baltimore, MD, December 2018.

“Distributional Semantic Methods for the Characterization Semantic and Pragmatic Facets of Language”. *National Research Council of Canada*. Ottawa, ON (remote presentation), February 2018.

“Mining Pregnancy-Related Health Information from Social Media”. *Department of Epidemiology, Biostatistics, and Informatics, Perelman School of Medicine, University of Pennsylvania*. Philadelphia, PA, November 2017.

“Detecting Gestation Period Using Social Media Data Analysis”. Poster Presentation at *The 16th Annual Biomedical Postdoctoral Research Symposium, University of Pennsylvania*. Philadelphia, PA, October 2017.

“Using Syntactic and Semantic Context to Explore Psychodemographic Differences in Self-reference”. Poster Presentation at *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Austin, TX, November 2016.

“Distributional Semantic Methods for Characterization of Atypical Language In Autism”. *Center for Language and Speech Processing, Johns Hopkins University*. Baltimore, MD, July 2016.

“Measuring idiosyncratic interests in children with autism spectrum disorder”. *The 4th Pacific Northwest NLP Workshop: NW-NLP*. Amazon, Seattle, WA, May 2016.

“Quantifying Restrictive and Repetitive Interest in Conversations of Autistic Children”. Poster Presentation at *4th Pacific Northwest NLP Workshop: NW-NLP*. Amazon, Seattle, WA, May 2016.

“Computational methods for categorization of disordered language”. *Positive Psychology Center, School of Arts and Sciences, University of Pennsylvania*. Philadelphia, PA, April 2016.

“Age and Gender Differences in Self-reference in Verb Categories”. *Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL)*. Philadelphia, PA, April 2016.

“Computational Semantic Analysis of Atypical Language in Autism”. *Department of Linguistics, Stony Brook University*. Stony Brook, NY, March 2016.

“Similarity Measures for Quantifying Restrictive and Repetitive Behavior in Conversations of Autistic Children”. *Computational Linguistics and Clinical Psychology Workshop (CLPsych) at NAACL*. Denver, CO, June 2015.

“Computational Semantic Analysis of Restrictive and Repetitive Behavior in Language Samples of Children with Autism”. Poster Presentation at *International Meeting for Autism Research*. Salt Lake City, UT, May 2015.

“Computational analysis of trajectories of linguistic development in autism”. Poster Presentation at *IEEE Spoken Language Technology Workshop (SLT)*. South Lake Tahoe, NV, December 2014.

“Detecting linguistic restricted interests in autism using distributional semantic models”. *Computational Linguistics & Clinical Psychology Workshop (CLPsych) at ACL*. Baltimore, MD, June 2014.

“Detecting linguistic restricted interests in autism”. *Center for Spoken Language Understanding, Oregon Health & Science University*. Portland, OR, June 2014.

“Children’s Differing Patterns of Discourse Marker Use in ASD and Typical Development”. Poster Presentation at *International Meeting for Autism Research (IMFAR)*. Atlanta, GA, May 2015.

“Distributional semantic applications in the evaluation of disordered speech”. *Center for Spoken Language Understanding, Oregon Health & Science University*. Portland, OR, May 2013.

“Distributional semantic models for the evaluation of disordered language”. Poster Presentation at *Annual Conference of the Institute on Development & Disability, Oregon Health & Science University*. Portland, OR, May 2013.

“Identifying Unexpected and Inappropriate Words in ASD Language Samples”. Poster Presentation at *Annual Conference of the Institute on Development & Disability, Oregon Health & Science University*. Portland, OR, May 2013.

“Distributional semantic applications in the evaluation of disordered speech”. *Center for Spoken Language Understanding, Oregon Health & Science University*. Portland, OR, February 2012.

“Knowledge Representation and Annotation for a Text-To-Scene System”. *Center for Spoken Language Understanding, Oregon Health & Science University*. Portland, OR, October 2012.

“Crowd-sourcing Semantic Information Collection”. *Language and Speech Processing Group, Department of Computer Science, Columbia University*. New York, NY, August 2012.

“Collecting Spatial Information for Locations in a Text-to-Scene Conversion System”. *2nd Pacific Northwest NLP Workshop: NW-NLP*. Microsoft Research. Redmond, WA, May 2012.

“Collecting Semantic Data by Mechanical Turk for the Lexical Knowledge Resource of a Text-to-Picture Generating System”. *Center for Spoken Language Understanding, Oregon Health & Science University*. Portland, OR, December 2010.

“Introduction to Lexical Resources and Ontologies”. *2nd Workshop on Natural Language Processing, Shahid Beheshti University*. Tehran, Iran, May 2007.

“Designing the WordNet for Persian Verbs”. *7th Conference on Iranian Linguistics, Allameh Tabatabaei University*. Tehran, Iran, December 2007.