Lu He

Department of Informatics
Donald Bren School of Information and Computer Science
University of California, Irvine

lu.he@uci.edu
www.luheholly.com

Research Interests

My research interests lie at the intersection of health informatics, human–computer interaction (HCI), and data science. My work has focused on developing more accurate and responsible computational systems using machine learning and natural language processing (NLP) to analyze both structured and unstructured health data. My research to date has contributed to evaluating and improving the performance of computational sentiment analysis, understanding the general public's opinions toward major health-related events and policies (e.g., mask and vaccine mandate during the COVID-19 pandemic), utilizing large-scale health data to inform better quality and efficiency of clinical care, and enhancing clinical understanding of rare diseases such as lymphoid malignancies.

Education

University of California, Irvine

- Ph.D. in Informatics, 2017–2023 (Expected)
- Advisor: Kai Zheng, PhD, FACMI

University of Minnesota, Twin Cities

- B.S. in Computer Science with Distinction, 2013-2017

Publications

Peer-Reviewed Journal Articles

J.9 Griffin A, **He L**, Sunjaya A, King A, Khan Z, Douthit B, Nwadiugw M, Subbin, V, Braunstein M, Nguyen V, Jaffe C, Schleyer T. Clinical, technical, and implementation characteristics of real-world health applications using FHIR. Journal of American Medical Informatics Association Open. 2022. (in press)

J.8 **He L**, Yin T, Zheng K. They may not work! An evaluation of eleven sentiment analysis tools on seven social media datasets. Journal of Biomedical Informatics. 2022;132:104142. PMID: 35835437

J.7 **He L**, He C. Help me DebunkThis: unpacking individual and community's collaborative work in information credibility assessment. Proceedings of the ACM on Human-Computer Interaction Volume 6, CSCW (November 2022). 2022. (in press)

J.6 Su Z, **He L**, Jariwala SP, Zheng K, Chen Y. "What is your envisioned future?": towards human-AI enrichment in data work of asthma care. Proceedings of the ACM on Human-Computer Interaction Volume 6, CSCW (November 2022). 2022. (in press)

J.5 He C, Liu H, **He L**, Lu T, Li B. More collaboration, less seriousness: Investigating new strategies for promoting youth engagement in government-generated videos during the COVID-19 pandemic in China. Computers in Human Behavior. 2021. DOI: 10.1016/j.chb.2021.107019

J.4 He C, **He L**, Lu T, Li B. Beyond entertainment: unpacking Danmaku and comments' role of information sharing and sentiment expression in online crisis videos. Proceedings of the ACM on Human-Computer Interaction Volume 5, CSCW (October 2021), 27 pages. DOI: 10.1145/3479555

J.3 **He L***, He C*, Reynolds TL, Bai Q, Huang Y, Li C, Zheng K, Chen Y. Why do people oppose mask wearing? A comprehensive analysis of US tweets during the COVID-19 pandemic. Journal of American Medical Informatics Association. (* equal contribution) 2021;28(7):1564–73. PMCID: PMC7989302 (**Editor's Choice and Featured Article**)

J.2 **He L**, Yin T, Hu Z, Chen Y, Hanauer DA, Zheng K. Developing a standardized protocol for computational sentiment analysis research using health-related social media data. Journal of American Medical Informatics Association. 2021;28(6):1125–34. PMCID: PMC8200276

J.1 Ma H, Smith C.E, **He L**, Narayanan S, Giaquinto R.A, Evans R, Hanson L, Yarosh S. Write for life: persisting in online health communities through expressive writing and social support. Proceedings of the ACM on Human-Computer Interaction Volume 1, CSCW (November 2017), 24 pages. DOI: 10.1145/3134708

Peer-Reviewed Full-Length Conference Papers

C.8 Guo Y, Zhu J, Huang Y, **He L**, He C, Li C, Zheng K. Public opinions toward COVID-19 vaccine mandates: a machine learning-based analysis of U.S. tweets. American Medical Informatics Association Annu Symp Proc. 2022. (in press) (**Student Paper Competition Finalist**)

C.7 **He L***, Song T*, Jiang Y, Yu P, Gong Y. To improve supportive care for patients taking oral anticancer agents. In: Proceedings of the 2021 World Congress on Health and Biomedical Informatic (MEDINFO 2021) 2022 Jun 6;290:547-551PMID: 35673076 (* equal contribution)

C.6 **He L**, He C, Wang Y, Hu Z, Zheng K, Chen Y. What do patients care about? Mining fine-grained patient concerns from online physician reviews through computer-assisted multi-level qualitative analysis. American Medical Informatics Association Annu Symp Proc. 2020;544–53. PMCID: PMC8075539 (**Student Paper Competition Finalist**)

C.5 **He L**, Zheng K. How do general-purpose sentiment analyzers perform when applied to health-related online social media data? Proceedings of the 2019 World Congress on Health and Biomedical Informatics (MEDINFO 2019). 2019;1208–12. PMCID: PMC8061710 (**Student Best Paper Nomination**)

C.4 Shehada ER, **He L**, Eikey EV, Jen M, Wong A, Young S, Zheng K. Characterizing frequent flyers of an emergency department using cluster analysis. In: Proceedings of the 2019 World Congress on Health and Biomedical Informatics (MEDINFO 2019). 2019;158–61. PMID: 31437905

C.3 Chi C, **He L**, Ravvaz K, Weissert J, P. Tonellato. Optimized decision support rules of precision warfarin treatment. 2018 Pacific Symposium on Biocomputing (PSB 2018). 2018;23:412-423. PMID: 29218901

C.2 Fan Y, **He L**, Zhang R. Evaluating automatic methods to extract patients' supplement use from clinical reports. IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2016, pp.1054-1061. DOI: 10.1109/BIBM.2016.7822668

C.1 Fan Y, **He L**, Pakhomov S, Melton G, Zhang R. Classifying supplement use status in clinical Notes. 2017 American Medical Informatics Association Joint Summits (AMIA 2017). 2017: 493–501. PMCID: PMC5543386 (**2nd Place in Student Paper Competition**)

Extended Abstracts, Workshops, and Posters

A.4 Griffin A, **He L**, Sunjaya A, King A, Khan Z, Douthit B, Nwadiugw M, Subbin, V, Braunstein M, Nguyen V, Jaffe C, Schleyer T. Assessment of real-world health applications on FHIR. Poster accepted to AMIA Annu Symp Proc. 2022.

A.3 **He L**, Cheng Y, Zhou T, Xian Y. Investigating the narratives of anti-Asian hate speech on Twitter during the COVID-19 pandemic. Podium Abstract, AMIA Annu Symp Proc. 2021

A.2 Goyal J, Ng DQ, Maddhuri J, Kumar AS, Jia S, **He L**, Wisseh C, Nguyen M, Lee J, McBane S, Zheng K, Hurley-Kim K, Nguyen Lee, Chan A, Cadiz CL. Predicting adverse drug events using the All of Us cohort data: A feasibility study. In: Proceedings of the 2021 American College of Clinical Pharmacy Annual Meeting. (ACCP 2021)

A.1 Yin T, **He L**. Challenges of applying sentiment analysis on health- related social media data. Southern California Natural Language Processing Symposium (SoCal 2019) (**Best Poster Award**)

Manuscripts in Progress

P.1 Ye J*, **He L***, Beestrum M. Implications for implementation and adoption of telehealth in low-and-middle income countries during the COVID-19 pandemic: systematic review of China's practices and experiences. (* indicates equal contribution) Major Revision and Resubmit in Journal of American Medical Informatics Association

P.2 **He L**, Ma H, Moldenhauer M, Zheng K. Extracting clinical and non-clinical information from clinical notes for veterans with lymphoid malignancies with limited expert annotation to assist clinical research. Podium Abstract Under Review

Honors and Awards

- Student Paper Competition Finalist, American Medical Informatics Association Annual Symposium (AMIA 2022) [C.8], 2022
- Graduate Dean's Dissertation Fellowship, University of California, Irvine, 2022
- Editor's Choice of Journal of American Medical Informatics Association (JAMIA) [J.3], 2021
- Student Paper Competition Finalist, American Medical Informatics Association Annual Symposium (AMIA 2020) [C.6], 2020
- Best Poster Award (1/40), Southern California Natural Language Processing Symposium (SoCal 2019) [A.1], 2019
- Student Best Paper Nomination, 2019 World Congress on Health and Biomedical Informatics (MedInfo '19) [C.2], 2019
- CRA-W Grad Cohort Workshop, Computing Research Association, 2018
- Graduate Dean Recruitment Fellowship, Department of Informatics, University of California, Irvine, 2017
- Dean's Award, Department of Informatics, University of California, Irvine, 2017
- 2nd Place in Student Paper Competition, 2017 American Medical Informatics Association Joint Summits (AMIA 2017) [C.1], 2017
- Undergraduate Special Recognition, Department of German, Nordic, Slavic and Dutch, University of Minnesota, 2014
- Maroon Global Excellence Scholarship (four-year), University of Minnesota, 2013-2017
- Dean's List (three-year), University of Minnesota, 2013-2016

Teaching Experience

University of California, Irvine

- Instructor of Record, ICS 33 (Intermediate Programming), 46 students
- Teaching Assistant, INF 151 (Project Management)

University of Minnesota, Twin Cities

- Teaching Assistant, CSci 1913 (Introduction to Data Structures and Algorithms)

Grant Activities

- UCI Internal (Ma, Zheng),2022-2023
  UCI ICTS/VA Long Beach Informatics Pilot Award, $50,000
  Outcomes from the National Database of Lymphoid Malignancies in US Veterans Using Data from the Veterans Health Administration, the Largest Integrated Healthcare System.
  Role: Grant writing and research assistant
- 2208383 (Bradley, Tang), 06/01/2022–05/31/2023
  NSF, $255,990
  STTR Phase I: An Integrated Biomedical Platform and Custom Algorithm to Optimize Feeding Protocols for Preterm Infants
  Role: Research assistant
- UCI Internal (Zheng), 06/01/2018–08/31/2018
  Academic Senate Council on Research, Computing and Libraries (CORCL), $3,500 (direct cost)
  Developing a Domain-Specific Sentiment Analyzer for Studying Health-Related Social Media Data
  Role: Grant writing and research assistant
- OT2OD026552 (Ohno-Machado, Anton-Culver), 04/01/2018–03/31/2023
  NIH Office of The Director, $58,453,967
  California Precision Medicine Research Program Consortium
  Role: Research assistant

Professional Services

Journal Editing

- JAMIA Student Editorial Board Member, 2022–2024

Program Committee

- 10th IEEE International Conference on Healthcare Informatics (ICHI), 2022

Reviewer

Journals

- Journal of Medical Internet Research (2019, 2020), BMC Medical Informatics and Decision Making (2020, 2021), Journal of Healthcare Informatics Research (2020, 2021), Journal of Biomedical Informatics (2021, 2022), IEEE Access (2022)

Conferences

- AMIA Annual Symposium (2018, 2019, 2020, 2021, 2022), AMIA Informatics Joint Summit (2019), CHI Late Breaking Work (2020), The Pacific Asia Conference on Information Systems (2021), CHI (2022)

Volunteer

- AMIA Symposium Student Volunteer (2017, 2021), AMIA Year-in-Review Student Working Group Volunteer (2020), Women in Academia Reading Group Organizer (2020)

Talks and Presentations

Invited Talks

- *"Characterizing Frequent Attenders of Emergency Department Using Cluster Analysis"*, Medical Intelligence and Innovation Institute (MI3), Children Hospital of Orange County (CHOC), Irvine, California, USA, April 8, 2019

Guest Lectures

- *"Computational Analysis of Social Media Data for Health Research: Applications, Challenges, and Opportunities"*, Center for Digital Health and Analytics, University of Texas Health School of Biomedical Informatics, Houston, Texas, USA, November 29, 2021. Host: Dr. Yang Gong
- *"Computational Analysis of Social Media Data for Health Research: Applications, Challenges, and Opportunities"*, Center for Digital Health and Analytics, University of Texas Health School of Biomedical Informatics, Houston, Texas, USA, October 21, 2021. Host: Dr. Sahiti Myneni

Conference Presentations

- *"Investigating the Narratives of Anti-Asian Hate Speech on Twitter During the COVID-19 Pandemic"*, AMIA Annual Symposium 2021, San Diego, California, USA. November 3, 2021
- *"What Do Patients Care About? Mining Fine-grained Patient Concerns from Online Physician Reviews Through Computer-Assisted Multi-level Qualitative Analysis"*, AMIA Annual Symposium 2020, Virtual Event (Recorded, Co-present with Changyang He), November 16, 2020
- *"What Do Patients Care About? Mining Fine-grained Patient Concerns from Online Physician Reviews Through Computer-Assisted Multi-level Qualitative Analysis"*, AMIA Annual Symposium 2020, Student Paper Competition (Virtual Live Presentation), November 15, 2020
- *"Characterizing Frequent Attenders of Emergency Department Using Cluster Analysis"*, MedInfo 2019, Lyon, France, August 25, 2019
- *"How Do General-purpose Sentiment Analyzers Perform on Health-related Social Media Data?"*, MedInfo 2019, Lyon, France, August 23, 2019

Mentoring

- Tasmima Khan (Undergraduate, UCI, 4/2021-11/2021)
- Tingjue Yin (Undergraduate, UCI, 3/2019-6/2020) [A1, J2, J8]
- Peilin Gan (Undergraduate, UCI, 1/2020-12/2020)
- Tianyang Zhou (Undergraduate, UCI, 3/2020-6/2020) [A3]
- Xinchen Zhang (Undergraduate, UCI, 1/2020-6/2020)
- Ya Cheng (Undergraduate, UCI, 3/2020-6/2020) [A3]
- Yongxu Xian (Undergraduate, UCI, 3/2020-6/2020) [A3]
- Daniel Davies (Undergraduate, UCI, 1/2019-6/2019)

- Su In Lee (Undergraduate, I-SURF Program, 6/2019-12/2019)
- Yiji Bae (Undergraduate, I-SURF Program, 6/2019-12/2019)
- Joohee Kwon (Undergraduate, I-SURF Program, 6/2019-12/2019)
- Haotian Hu (Undergraduate, UCI, 10/2019-12/2019)

Media Coverage

- KXAN News: Do face masks work? Here are 49 scientific studies that explain why they do, December, 2021
- Women in AMIA: AMIA as a Catalyst for Collaboration, August, 2021

Skills

Research Methods

Qualitative Methods

- Interview, observational study, user study, grounded theory, wizard-of-oz

Quantitative Methods

- Survey, applied machine learning and deep learning, natural language processing, log analysis, social network analysis, statistical analysis (hypothesis testing, linear models, generalized linear models, longitudinal data analysis, linear mixed effect models, generalized estimating equation, survival analysis)

Programming Languages

- Python (proficient), R (medium), Matlab (medium), Java (medium), C/C++ (medium), SQL (medium), Lisp (familiar), Clojure (familiar), Go (familiar), Ruby (familiar)

Frameworks and Packages

- Python (scikit-learn, Numpy, Pandas, NLTK, Spacy, Gensim, Keras, PyTorch, Django, Tensorflow), R (ggplot, tidyverse, tidytext, dplyr, cluster, mltools, sna, igraph)

Databases

- MySQL, PostgreSQL

Languages

- Mandarin Chinese (Native), English (Full professional proficiency), Japanese (JLPT N3), German (Limited proficiency)

Last Updated: September 2022