

Junjie Luo

Biomedical Informatics & Data Science
School of Medicine
Johns Hopkins University

jluo41@jhu.edu
+1 240 524 3597

EDUCATION

- Ph.D. Health Science Informatics, Biomedical Informatics & Data Science, Johns Hopkins School of Medicine, 2022-2026 (expected)
- Ph.D. Information Systems, R.H. Smith School of Business, University of Maryland, College Park, 2020-2022 (transferred)
- M.S. Data Science, School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, 2018-2020
- B.B.A. Applied Economics, School of Management and Economics *summa cum laude*, The Chinese University of Hong Kong, Shenzhen, 2014-2018

RESEARCH INTERESTS

My research is focused on designing and developing healthcare AI using real-world data, with special emphasis on human factors and behaviors to enhance health and wellness.

Artificial Intelligence: natural language processing, LLM, human-centered AI, foundation large models for healthcare, reinforcement learning.

Healthcare: big data in healthcare, health IT, EHR, mobile health, causal models, multi-modal sensor data modeling.

Behavioral Economics: human behavior, behavioral bias, social disparities, and experimental economics.

WORKING PAPERS

- 2025 **Junjie Luo**, Mansur Shomali, Abhimanyu Kumbara, Anand Iyer, Rui Han, Ritu Agarwal, Guodong (Gordon) Gao. “Let Curves Speak: A Continuous Glucose Monitor based Large Sensor Foundation Model for Diabetes Management” working paper, (2025).
- 2025 **Junjie Luo**, Di Hu, Rui Han, Diyang Lyu, Ritu Agarwal, Jehan El-Bayoumi, Guodong (Gordon) Gao, Nawar Shara. “Risk-Based Prediabetes Onset Subtypes Using EHR Data: a Machine Learning Approach” working paper, (2025).

PUBLICATIONS

- 2024 Mansur Shomali, **Junjie Luo**, Abhimanyu Kumbara, Anand Iyer, Guodong (Gordon) Gao. “CGM-GPT: A Transformer Based Glucose Prediction Model to Predict Glucose Trajectories at Different Time Horizons” *Machine Learning for Healthcare*. (2024).
- 2022 Weiguang Wang, **Junjie Luo**, Michelle Dugas, Guodong (Gordon) Gao, Ritu Agarwal, Rachel Werner. “Recency in Online Physician Ratings” *JAMA Internal Medicine*, (2022).

- 2021 Jichao Sun, Yefeng Zheng, Wenhua Liang, Zifeng Yang, Zhiqi Zeng, Tiegang Li, **Junjie Luo**, Man Tat Alexander Ng, Jianxing He, Nanshan Zhong. “Quantifying the Effect of Public Activity Intervention Policies on COVID-19 Pandemic Containment Using Epidemiologic Data From 145 Countries” *Value in Health*, (2021).
- 2021 Ritu Agarwal, Michelle Dugas, Jui Ramaprasad, **Junjie Luo**, Gujie Li, and Guodong (Gordon) Gao. “Socioeconomic privilege and political ideology are associated with racial disparity in COVID-19 vaccination.” *Proceedings of the National Academy of Sciences* 118, no. 33 (2021).
- 2021 **Junjie Luo**, Xi Chen, Jichao Sun, Yuejia Xiang, Ningyu Zhang and Xiang Wan. “Field Embedding: A Unified Grain Based Framework for Word Representation” *North American Association for Computational Linguistics*, (2021).

CONFERENCE PRESENTATIONS

- 2024.5 **Junjie Luo**, Abhimanyu Kumbara, Anand Lyer, Mansur Shomali and Guodong (Gordon) Gao. “Using an Automated, Real-time Data Enabled Feature Engineering Process to Predict Future Weight Outcomes”. *Conference on Health IT and Analytics’2024*.
- 2023.7 **Junjie Luo**, Gordon Gao, Ritu Agarwal. “Assessing Dataset Quality to AI Bias and Healthcare Disparity: A Study on EHR-Based Prediabetes to Diabetes Risk Prediction Models”. *INFORMS Healthcare’2023*.
- 2023.7 **Junjie Luo**, Mansur Shomali, Abhimanyu Kumbara, Anand Lyer, and Guodong (Gordon) Gao. “Using Early Engagement Data from a Digital Health to Predict Future Health Outcomes”. *INFORMS Healthcare’2023*.
- 2023.5 **Junjie Luo**, Aishwarya Shukla, Jie Mein Goh, Guodong (Gordon) Gao, Ritu Agarwal. “Relationships Between Patient Satisfaction and Physician Overprescription: Evidence from Online Reviews and CMS Claim Data”. *Production and Operations Management Society’2023*.
- 2023.5 **Junjie Luo**, Abhimanyu Kumbara, Anand Lyer, Mansur Shomali, and Guodong (Gordon) Gao. “Using Early Engagement and CGM Data from a Digital Health Solution to Predict Future Engagement Patterns”. *Conference on Health IT and Analytics’2023*.
- 2022.11 **Junjie Luo**, Casey Kindall, Sara-Laure Faraji, Guodong (Gordon) Gao, Ritu Agarwal, Sally Simpson. “Foretelling the Physician’s Future: Healthcare Fraud Prediction with Behavioral Big Data”. *American Society of Criminology Annual Meeting’2022*.
- 2022.6 **Junjie Luo**, Guodong (Gordon) Gao, Ritu Agarwal, and Sally Simpson. ““Minority Report” Comes True: Using Behavioral Big Data to Predict Physician Fraud”. *Academy Health’2022*.
- 2022.4 **Junjie Luo**, Shiping Liu, Guodong (Gordon) Gao, Ritu Agarwal. “Racial Disparity in the Availability of Social Determinants of Health Information in EHR Data”. *Production and Operations Management Society’2022*.
- 2022.3 **Junjie Luo**, Guodong (Gordon) Gao, Ritu Agarwal, and Sally Simpson. ““Minority Report” Comes True: Using Behavioral Big Data to Predict Physician Fraud”. *Conference on Health IT and Analytics’2022*.
- 2021.6 **Junjie Luo**, Xi Chen, Jichao Sun, Yuejia Xiang, Ningyu Zhang and Xiang Wan. “Field Embedding: A Unified Grain Based Framework for Word Representation”. *North American Association for Computational Linguistics’2021*.

TEACHING

Johns Hopkins University

2025 spring AI Essentials for Business, Lecturer
2024 fall Clinical Data Analysis with Python, Teaching Assistant.
2024 fall AI Essentials for Business, AI tech lead.
2024 spring Big Data Machine Learning, AI tech lead.

University of Maryland, College Park

2021 summer AI Leadership in Healthcare Exec Ed program, Head of AI Specialists.

The Chinese University of Hong Kong, Shenzhen

2017 spring DMS2051: IT in Business Application, Teaching Assistant.

EMPLOYMENT EXPERIENCE

2020 Algorithm Engineer, Jarvis AI Lab, Healthcare Group, Tencent Ltd, Shenzhen, China
2017-20 Research Assistant, Shenzhen Research Institute of Big Data, Shenzhen, China
2017 Data Mining Engineer, Ping An of China, Ping An Technology Co. Ltd, Shenzhen, China
2016-17 Research Assistant, School of Management and Economics, CUHK, Shenzhen

INDUSTRY COLLABORATION

DrFirst Inc., WellDoc Inc., MedStar Health, Inovalon Inc., Children's National Hospital.

RESEARCH EXPERIENCE

Clinical Text Processing: Using cutting-edge natural language processing and deep learning methods to learn the representation of clinical notes (both English and Chinese). The methods includes BERT-like pretrained language models, linguistic fields representation, and knowledge graph embeddings.

Physician-Patient Behavior Bias Study with EHR Data: Utilizing MedStar's voluminous EHR data, we discover behavioral biases and cognitive biases of both physicians and patients.

Facebook Ad Campaign for Health Awareness: Using experimental methods to increase the health awareness related to HIV prevention of adolescent girls and young women in South Africa.

WellDoc CGM Analysis: Using Continuous Glucose Monitor (CGM) data to study the behavior patterns of Type-II Diabetes patients and develop methods to increase their healthy behaviors.

Call Center Hourly-Based AI Recommendation: Deploying AI and behavioral big data to study optimal timing for patient call-backs.

Racial Disparity in COVID-19 Vaccination: Using uniquely curated public data to reveal the factors that are associated with racial disparity in COVID-19 vaccination rate.

Online Physician Reviews: Collecting all online reviews for one million MD doctors in the US, and studying the recency of the online reviews.

HOURS AND AWARDS

2020	R.H.Smith School Graduate Assistant Scholarship
2020	Academic Performance Scholarship
2018	Admission Scholarship for Masters Program
2015-18	Dean's List Honor
2016-17	Academic Performance Scholarship
2016-17	Undergraduate Research Award
2015-17	National Scholarship
2014	Admission Scholarship for Undergraduate Program

TOOLKITS DEVELOPED

recfldtkn: Conceptualizing human subjects as collections of time-stamped records with multi-modal values, this pipeline package transforms these records into AI datasets and trains various types of AI models. This versatile pipeline is designed for both training and inference, allowing for implementation on real-world business platforms.

cctner: A model that automatically extracts body, symptom, disease, examination, and other clinical entities from Chinese clinical texts.

ATol: A website-based and AI-supported annotation tool for collaborative annotation.

nlptext: A toolkit for preprocessing large English and Chinese corpus data with annotation labels and linguistic field grain sequences.

SKILLS

Programming: Python, R, MATLAB, SQL, STATA, C/C++.

Analytics: PyTorch, Tensorflow, Speech Recognition, Natural Language Processing, Language Language Model, Multi-Modal Sensor Modeling, Topic Modeling, Sentimental Analysis, Stance Detection, ArcGIS, Tableau, Web Scraping, Web Development.

MLOps: AWS, Hadoop, Spark, Hive, Docker, SageMaker, Kubernetes, EC2.

MEMBERSHIPS

Association for Computational Linguistics

American Medical Informatics Association

AcademyHealth

INFORMS

Production and Operations Management Society