

ZONGHAI YAO

+1 (413) 824-5685 ◊ zonghaiyao@umass.edu ◊ [Homepage](#) ◊ [Google Scholar](#) ◊ [Linkedin](#)

EDUCATION

UMass Amherst	Ph.D. in Computer Science	Amherst, MA	Sep.2021 - May.2026
UMass Amherst	M.S. in Computer Science	Amherst, MA	Sep.2019 - May.2021
Nankai University	B.S. in Computer Science	Tianjin, China	Sep.2015 - Jun.2019

AWARDS

- Outstanding Research Award Center for Biomedical and Health Research in Data Science Aug 2025
- UMass Amherst Outstanding PhD Portfolio Award Dec 2024

AI/NLP & BIOMEDICAL INFORMATICS CONFERENCES

- **ChatCLIDS: Simulating Persuasive AI Dialogues to Promote Closed-Loop Insulin Adoption in Type 1 Diabetes Care** (* indicates equal contribution)
Z. Yao*, T. Chafekar*, J. Wang, S. Han, F. Ouyang, J. Qian, L. Li, H. Yu AAAI-26 AI for Social Impact
- **PRIME: Planning and Retrieval-Integrated Memory for Enhanced Reasoning**
H. Tran, Z. Yao, NL. Tran, Z. Yang, F. Ouyang, S. Han, R. Rahimi, H. Yu AAAI 2026
- **DischargeSim: A Simulation Benchmark for Educational Doctor-Patient Communication at Discharge**
Zonghai Yao*, Michael Sun*, Won Seok Jang, Sunjae Kwon, Soie Kwon, and Hong Yu EMNLP 2025
- **From Scores to Steps: Diagnosing and Improving LLM Performance in Evidence-Based Medical Calculations** (+ indicates Co-corresponding)
B. Wang*, I. Xia*, Y. Zhang, J. Wang, F. Ouyang, S. Han, H. Yu+, and Z. Yao+ EMNLP 2025 (oral)
- **MedCOD: Enhancing Medical Translation of LLMs Using Enriched Chain-of-Dictionary Framework**
MS. Salim, L. Fu, AA. Ramakrishnan, Z. Yao+ and H. Yu+ EMNLP Findings 2025
- **Chatbot To Help Patients Understand Their Health**
WS. Jang*, H. Tran*, M. Mistry, S. Gandluri, Y. Zhang, S. Sultana, S. Kwon, Z. Yao+, H. Yu+ EMNLP Findings 2025
- **RARE: Retrieval-Augmented Reasoning Enhancement for Large Language Models**
Hieu Tran, Zonghai Yao, Junda Wang, Yifan Zhang, Zhichao Yang, and Hong Yu ACL 2025
- **MCQG-SRefine: Multiple Choice Question Generation and Evaluation with Iterative Self-Critique, Correction, and Comparison Feedback**
Z Yao*, A Parashar*, H Zhou, WS Jang, F Ouyang, Z Yang, and H Yu NAACL 2025 (oral)
- **SYNFAC-EDIT: Synthetic Imitation Edit Feedback for Factual Alignment in Clinical Summarization**
P. Mishra*, Z. Yao*, P. Vashisht, F. Ouyang, B. Wang, VD. Mody, and H. Yu EMNLP 2024
- **README: Bridging Medical Jargon and Lay Understanding for Patient Education through Data-Centric NLP**
Z. Yao, N. Kantu, G. Wei, H. Tran, Z. Duan, S. Kwon, Z. Yang, and H. Yu EMNLP Findings 2024
- **Large Language Models are In-context Teachers for Knowledge Reasoning**
Jiachen Zhao, Zonghai Yao, Zhichao Yang, and Hong Yu EMNLP Findings 2024
- **NoteChat: A Dataset of Synthetic Doctor-Patient Conversations Conditioned on Clinical Notes**
J. Wang*, Z. Yao*, Z. Yang, H. Zhou, R. Li, X. Wang, Y. Xu, and H. Yu ACL Findings 2024
- **LocalTweets to LocalHealth: A Mental Health Surveillance Framework Based on Twitter Data**
V. Deshpande, M. Lee, Z. Yao, Z. Zhang, J. Gibbons, and H. Yu COLING 2024
- **Do Physicians Know How to Prompt? The Need for Automatic Prompt Optimization Help in Clinical Note Generation**
Zonghai Yao, Ahmed Jaafar, Beining Wang, Zhichao Yang, Hong Yu BioNLP 2024
- **Improving Summarization with Human Edits**
Zonghai Yao, Benjamin J Schloss, and Sai P. Selvaraj EMNLP 2023

- **Revisiting the Architectures like Pointer Networks to Efficiently Improve the Next Word Distribution, Summarization Factuality, and Beyond**
H. Chang*, **Z. Yao***, A. Gon, H. Yu, and A. McCallum ACL Findings 2023
- **Multi-label Few-shot ICD Coding as Autoregressive Generation with Prompt**
Zhichao Yang, Sunjae Kwon, **Zonghai Yao**, and Hong Yu AAAI 2023
- **Can LLMs generate high-quality synthetic note-oriented doctor-patient conversations?**
Junda Wang*, **Zonghai Yao***, Avijit Mitra, Samuel Osebe, Zhichao Yang and Hong Yu ClinicalNLP 2023
- **Context Variance Evaluation of Pretrained Language Models for Prompt-based Biomedical Knowledge Probing**
Z. Yao, Y. Cao, Z. Yang, and H. Yu AMIA 2023 (oral)
- **MedJEx: A Medical Jargon Extraction Model with Wiki's Hyperlink Span and Contextualized Masked Language Model Score**
S. Kwon, **Z. Yao**, H. Jordan, D. Levy, B. Corner, and H. Yu EMNLP 2022
- **Extracting Biomedical Factual Knowledge Using Pretrained Language Model and Electronic Health Record Context**
Z. Yao, Y. Cao, Z. Yang, V. Deshpande, and H. Yu AMIA 2022 (oral)
- **Improving Formality Style Transfer with Context-Aware Rule Injection**
Zonghai Yao and Yu Hong ACL 2021
- **Zero-shot Entity Linking with Efficient Long Range Sequence Modeling**
Zonghai Yao, Liangliang Cao and Huapu Pan EMNLP Findings 2020

BIOMEDICAL INFORMATICS & COMPUTATIONAL HEALTH JOURNALS

- **Automated identification of eviction status from electronic health record notes**
Z. Yao, J. Tsai, W. Liu, D. Levy, E. Druhl, J. Reisman, and H. Yu
Journal of the American Medical Informatics Association, Volume 30, Issue 8, August 2023, Pages 1429-1437
- **Unveiling GPT-4V's hidden challenges behind high accuracy on USMLE questions: Observational Study**
Z. Yang*, **Z. Yao***, M. Tasmin, P. Vashisht, W. Jang, B. Wang, D. Berlowitz, and H. Yu
Journal of Medical Internet Research, 27, e65146.
- **PaniniQA: Enhancing Patient Education Through Interactive Question Answering**
P. Cai*, **Z. Yao***, F. Liu, D. Wang, M. Reilly, H. Zhou, L. Li, Y. Cao, A. Kapoor, A. Bajracharya, D. Berlowitz and H. Yu
Transactions of the Association for Computational Linguistics (2023) 11: 1518-1536. (TACL)
- **BioInstruct: instruction tuning of large language models for biomedical natural language processing**
Hieu Tran, Zhichao Yang, **Zonghai Yao**, and Hong Yu
Journal of the American Medical Informatics Association, Volume 31, Issue 9, September 2024, Pages 1821-1832
- **Development of a Surveillance System to Identify Incidence of Evictions Among Patients in Veterans Affairs Medical Centers Across the United States**
J. Tsai, SS. Rajan, **Z. Yao**, J. Reisman, W. Liu, and H. Yu
Journal of Community Health, 1-10.
- **Association between PTSD and health-related social needs in US Veterans: an NLP analysis using Veterans Health Administration Data**
F. Ouyang, W. Hu, J. Reisman, T. Pogoda, K. Carlson, W. Liu, Y. Moradi, Y. Zhang, S. Han, S. Sultana, W. Jang, **Z. Yao**, A. Mitra, H. Yu
Journal of Affective Disorders, 120643.
- **Named Entity Location Prediction Combining Twitter and Web**
Y. Liu, W. Shen, **Z. Yao**, J. Wang, Z. Yang and X. Yuan
IEEE Transactions on Knowledge and Data Engineering, Volume 33, Issue 11, 01, November 2021, Pages 3618-3633

UNDER REVIEW (PREPRINT)

- **Medical thinking with multiple images**
Z. Yao*, B. Wang*, Y. Zhang, J. Wang, I. Xia, Z. Tang, S. Han, F. Ouyang, Z. Yang, A. Cohan, H. Yu [link](#)

- **MedQA-CS: Benchmarking Large Language Models Clinical Skills Using an AI-SCE Framework**
Z. Yao, Z. Zhang, C. Tang, X. Bian, Y. Zhao, Z. Yang, J. Wang, H. Zhou, W. Jang, F. Ouyang, H. Yu [link](#)
- **A Survey on LLM-based Multi-Agent AI Hospital**
Zonghai Yao and Hong Yu [link](#)
- **SynthEHR-Eviction: Enhancing Eviction SDoH Detection with LLM-Augmented Synthetic EHR Data**
Z. Yao*, Y. Zhao*, A. Mitra, D. Levy, E. Druhl, J. Tsai, and H. Yu [link](#)
- **ChatThero: A Language Agent for Recovery Support**
J. Wang*, Z. Yao*, Z. Yang, J. Qian, L. Li, H. Yu [link](#)
- **Exploiting Tree Structure for Credit Assignment in RL Training of LLMs**
H. Tran*, Z. Yao*, H. Yu [link](#)
- **RiTeK: A Dataset for Large Language Models Complex Reasoning over Textual Knowledge Graphs**
J. Huang*, M. Li*, Z. Yao, D. Li, Y. Zhang, Z. Yang, Y. Xiao, F. Ouyang, X. Li, S. Han, H. Yu [link](#)
- **Iterative Critique-Refine Framework for Enhancing LLM Personalization**
D. Maram, D. Gandhi, Z. Yao, G. Akkinapalli, F. Dernoncourt, Y. Wang, R. Rossi, N. Ahmed [link](#)
- **JMLR: Joint Medical LLM and Retrieval Training for Enhancing Reasoning and Professional QA**
Junda Wang, Zhichao Yang, Zonghai Yao+, and Hong Yu+ [link](#)
- **Personalizing medical text generation with readability-controlled instruction learning**
H. Tran*, Z. Yao*, W. Jang, S. Sultana, A. Chang., Y. Zhang, and H. Yu [link](#)
- **Enhancing LLMs for Identifying and Prioritizing Important Medical Jargons from Electronic Health Record Notes Utilizing Data Augmentation**
W. Jang*, S. Sultana*, Z. Yao*, H. Tran, Z. Yang, S. Kwon, and H. Yu [link](#)

SERVICE

Reviewing

- Area Chair: NAACL ARR (2025), ACL ARR (2025), EMNLP ARR (2025), AACL ARR (2025), EACL ARR (2025)
- Conference Reviewer: [ICLR](#), [NeurIPS](#), [ICML](#), [AAAI](#), [ACL](#), [EMNLP](#), [NAACL](#), [COLING](#), [EACL](#) *2023 - Present*
- Medical Journal Reviewer (30+ reviews) *2022 - Present*
 - [npj Digital Medicine](#)
 - [npj Health Systems](#)
 - [Scientific Reports](#)
 - [Journal of Medical Internet Research](#)
 - [Journal of the American Medical Informatics Association](#)
 - [Bioinformatics](#)
 - [Expert Systems With Applications](#)
 - [BMC Medical Informatics and Decision Making](#)
 - [BMC Medical Research Methodology](#)
 - [European Heart Journal - Digital Health](#)

Organizing

- Session Chairs for AACL (2025)
- Organizer for UMass BioNLP Workshop *Dec 2023, 2024, 2025*

Committees

- Committee for American Medical Informatics Association Student Working Group [link](#) *Jan 2025 - Present*

TEACHING

- **Guest Lecturer** - UMass Lowell - COMP.5770 Methods in Data Science *Spring 2025*
- **Guest Lecturer** - UMass Lowell - COMP.5800 Topics in Computer Science *Fall 2024*

- **Teaching Assistant** - UMass Amherst - COMPSCI 240: Reasoning Under Uncertainty *Spring 2024*
- **Guest Lecturer** - UMass Lowell - COMP.5800 Topics in Computer Science *Fall 2023*

MENTORING

- **Zhenyu Zhang** - MS Student (Stanford) *Fall 2025-Present*
- **Nhat Minh Le** - Undergrad (UMass Amherst) *Fall 2025-Present*
- **Lucas Lu** - MS Student (UMass Amherst) *Fall 2025-Present*
- **Benlu Wang** - MS Student (Yale) *Spring 2025-Present*
- **Iris Xia** - MS Student (Yale) *Spring 2025-Present*
- **Sravanthi Machcha** - MS Student (UMass Amherst) *Spring 2025-Present*
- **Lianne Fu** - MS Student (UMass Amherst) *Spring 2025-Present*
- **Talha Chafekar** - MS Student (UMass Amherst) *Fall 2024-Present*
- **Sushrita Yerra** - MS Student (UMass Amherst) - Amazon *Fall 2024-Present*
- **Parth Vashisht** - MS Student (UMass Amherst) - Ph.D. Student (UMass Lowell) *Fall 2023-Present*
- **Adiba Haque** - MS Student (UMass Amherst) - Ph.D. Student (UMass Lowell) *Fall 2023-Present*
- **Michael Sun** - MS Student (UMass Amherst) - Athenahealth *Fall 2024-Fall 2025*
- **Chaolong Tang** - Undergrad&MS Student (UMass Amherst) *Spring 2024-Spring 2025*
- **Youxia Zhao** - MS Student (UMass Amherst) *Spring 2024-Spring 2025*
- **Zihao Zhang** - MS Student (UMass Amherst) - Ph.D. Student (Emory) *Spring 2024-Spring 2025*
- **Xingyu Bian** - Undergrad&MS Student (UMass Amherst) *Spring 2024-Spring 2025*
- **Aditya Parashar** - MS Student (UMass Amherst) - Capital One *Spring 2024-Spring 2025*
- **Prakamy Mishra** - MS Student (UMass Amherst) - AMD *Fall 2023-Fall 2024*
- **Jiachen Zhao** - MS Student (UMass Amherst) - Ph.D. Student (Northeastern) *Fall 2023-Spring 2024*
- **Nandyala Siddharth Kantu** - MS Student (UMass Amherst) *Fall 2023-Spring 2024*
- **Ahmed Jaafar** - MS Student (UMass Amherst) *Fall 2023-Spring 2024*
- **Minhwa Lee** - MS Student (UMass Amherst) - Mayo Clinic *Fall 2022-Spring 2024*
- **Lingxi Li** - Undergrad&MS Student (UMass Amherst) - Contextual AI *Fall 2022-Spring 2024*
- **Yi Cao** - Undergrad&MS Student (UMass Amherst) - Amazon *Fall 2021-Spring 2023*
- **Guanghao Wei** - Undergrad (UMass Amherst) - MS Student (Cornell) *Fall 2021-Spring 2023*

INVITED TALKS

- **Chatbot for Behavior Change and Therapeutic Support in Addiction Recovery**
Veteran Affairs *August 2025*
- **Bridging Health Literacy Gaps: Improving Patient Understanding of Electronic Health Records**
Zhongshan Ophthalmic Center of Sun Yat-sen University *June 2024*
- **Helping Veterans Understand Clinical Notes with Advanced AI Systems**
Veteran Affairs *May 2024*