Zifeng Wang

Department of Computer Science
University of Illinois Urbana Champaign

zifengw2@illinois.com
https://zifeng.wang
github.com/RyanWangZf

EDUCATION

Ph.D. Computer Science, University of Illinois Urbana Champaign, Urbana, USA, 2021-present

M.S. Data Science, Tsinghua University, Shenzhen, China, 2018-2021

B.S. Civil Engineering, Tongji University, Shanghai, China, 2014-2018

APPOINTMENTS

2021– University of Illinois Urbana Champaign
Research Assistant, Supervisor: Jimeng Sun

2023 Amazon Web Service
Applied Scientist Intern, Mentor: Zichen Wang

2023 Medidata, Dassault Systèmes
Research Intern, Mentor: Mandis Beigi

2022 Amplitude
Research Intern, Mentor: Cao Xiao

2019-2021 Tencent
Research Intern, Mentor: Yefeng Zheng

RESEARCH AREAS

The development of foundation models, including large language models (LLMs), tabular models, vision-language models, and biomedical multimodal models, for the entire clinical trial process.

My recent focus has shifted towards creating LLM-driven systems to enhance human-AI collaboration in various clinical trial tasks, including systematic review, trial design, and clinical data science.

SELECTED PAPERS

2024 **Wang, Z.**, Cao, L., Danek, B., Zhang, Y., Jin, Q., Lu, Z., Sun, J., "Accelerating Clinical Evidence Synthesis with Large Language Models." In: *arXiv preprint arXiv:2406.17755* (2024)

2024 Lin, J., Xu, H., **Wang, Z.**, Wang, S., Sun, J., "Panacea: A foundation model for clinical trial search, summarization, design, and recruitment." In: *medRxiv* (2024)

2024 Jiang, P., Xiao, C., **Wang, Z.**, Bhatia, P., Sun, J., Han, J., "TriSum: Learning Summarization Ability from Large Language Models with Structured Rationale." In: *NAACL'24*. 2024, pp. 2805–2819

2024 Jiang, P., Lin, J., **Wang, Z.**, Sun, J., Han, J., "GenRES: Rethinking Evaluation for Generative Relation Extraction in the Era of Large Language Models." In: *NAACL'24*. 2024, pp. 2820–2837

2024 Cao, L., **Wang, Z.**, Xiao, C., Sun, J., "PILOT: Legal Case Outcome Prediction with Case Law." In: *NAACL'24*. 2024, pp. 609–621

2023 Wang, R., **Wang, Z.**, Sun, J., "Unipredict: Large language models are universal tabular predictors." In: *arXiv preprint arXiv:2310.03266* (2023)

2023 **Wang, Z.**, Wang, Z., Srinivasan, B., Ioannidis, V. N., Rangwala, H., Anubhai, R., "BioBridge: Bridging Biomedical Foundation Models via Knowledge Graphs." In: *ICLR'24*. 2023

2023 Feng, T., **Wang, Z.**, Sun, J., "Citing: Large language models create curriculum for instruction tuning." In: *arXiv preprint arXiv:2310.02527* (2023)

2023 Wen, Y., **Wang, Z.**, Sun, J., "Mindmap: Knowledge graph prompting sparks graph of thoughts in large language models." In: *ACL'24*. 2023

2023 Das, T., **Wang, Z.**, Sun, J., "Twin: Personalized clinical trial digital twin generation." In: *KDD'23*. 2023, pp. 402–413

2023 Jin, Q., **Wang, Z.**, Floudas, C. S., Chen, F., Gong, C., Bracken-Clarke, D., Xue, E., Yang, Y., Sun, J., Lu, Z., "Matching Patients to Clinical Trials with Large Language Models." In: *ArXiv* (2024), arXiv–2307

2023 **Wang, Z.**, Theodorou, B., Fu, T., Xiao, C., Sun, J., "Pytrial: Machine learning software and benchmark for clinical trial applications." In: *arXiv preprint arXiv:2306.04018* (2023)

2023 **Wang, Z.**, Gao, C., Xiao, C., Sun, J., "MediTab: Scaling Medical Tabular Data Predictors via Data Consolidation, Enrichment, and Refinement." In: *IJCAI'24*. 2023

2023 **Wang, Z.**, Xiao, C., Sun, J., "AutoTrial: Prompting Language Models for Clinical Trial Design." In: *EMNLP'23*. 2023, pp. 12461–12472

2023 **Wang, Z.**, Xiao, C., Sun, J., "SPOT: sequential predictive modeling of clinical trial outcome with meta-learning." In: *ACM-BCB'23*. 2023, pp. 1–11

2022 **Wang, Z.**, Sun, J., "Transtab: Learning transferable tabular transformers across tables." In: *NeurIPS'22* 35 (2022), pp. 2902–2915

2022 **Wang, Z.**, Wu, Z., Agarwal, D., Sun, J., "MedCLIP: Contrastive Learning from Unpaired Medical Images and Text." In: *EMNLP'22*. 2022

2022 **Wang, Z.**, Sun, J., "PromptEHR: Conditional Electronic Healthcare Records Generation with Prompt Learning." In: *EMNLP'22*. 2022, pp. 2873–2885

2022 **Wang, Z.**, Sun, J., "Trial2Vec: Zero-Shot Clinical Trial Document Similarity Search using Self-Supervision." In: *Findings of EMNLP'22*. 2022

2022 **Wang, Z.**, Sun, J., "Survtrace: Transformers for survival analysis with competing events." In: *ACM-BCB'22*. 2022, pp. 1–9

2021 **Wang, Z.**, Wen, R., Chen, X., Huang, S.-L., Zhang, N., Zheng, Y., "Finding Influential Instances for Distantly Supervised Relation Extraction." In: *COLING'22*. 2022, pp. 2639–2650

2021 **Wang, Z.**, Huang, S. L., Kuruoglu, E. E., Sun, J., Chen, X., Zheng, Y., "PAC-Bayes Information Bottleneck." In: *ICLR'22*. 2022

2021 **Wang, Z.**, Yang, Y., Wen, R., Chen, X., Huang, S.-L., Zheng, Y., "Lifelong learning based

disease diagnosis on clinical notes." In: *PAKDD'21*. Springer. 2021, pp. 213–224

2020 **Wang, Z.**, Wen, R., Chen, X., Cao, S., Huang, S.-L., Qian, B., Zheng, Y., "Online disease diagnosis with inductive heterogeneous graph convolutional networks." In: *WWW'21*. 2021, pp. 3349–3358

2020 **Wang, Z.**, Chen, X., Wen, R., Huang, S.-L., Kuruoglu, E., Zheng, Y., "Information theoretic counterfactual learning from missing-not-at-random feedback." In: *NeurIPS'20* 33 (2020), pp. 1854–1864

2019 **Wang, Z.**, Zhu, H., Dong, Z., He, X., Huang, S.-L., "Less is better: Unweighted data subsampling via influence function." In: *AAAI'20*. Vol. 34. 04. 2020, pp. 6340–6347

2019 **Wang, Z.**, Zhang, Y., Mosalam, K. M., Gao, Y., Huang, S.-L., "Deep semantic segmentation for visual understanding on construction sites." In: *Computer-Aided Civil and Infrastructure Engineering* 37.2 (2022), pp. 145–162

2018 **Wang, Z.**, Li, S., "Data-driven risk assessment on urban pipeline network based on a cluster model." In: *Reliability Engineering & System Safety* 196 (2020), p. 106781

INVITED TALKS

2024 Bridging biomedical foundation models, invited by Xuegong Lab@Tsinghua University.

2023 Bridging biomedical foundation models, invited by AI4Science@ByteDance.

2023 Automate clinical trial design with large language models, invited by Medidata.

2023 Medical vision-language modeling, invited by University of Toronto.

2023 Medical vision-language modeling, invited by MedAI@Stanford University.

2023 Medical vision-language modeling, invited by NCBI@NIH.

2023 Unifying large language models and knowledge graphs, invited by Titan Lab@AWS.

MEDIA COVERAGE

2024 Connecting Patients to Clinical Trials With Artificial Intelligence, American Urological Association (AUA News).

2024 How AI is being used to accelerate clinical trials , Nature.

AWARDS

2023 Yee Memorial Fellowship, UIUC

2022 NeurIPS Scholar Award, NeurIPS

2022 Yunni & Maxine Pao Memorial Fellowship, UIUC

2021 Best Student Paper Award, PAKDD

SERVICE

Academic Journal and Conference Peer Review

Program Committee: NeurIPS (2022, 2023, 2024), EMNLP (2022, 2023), IJCAI (2022, 2023, 2024), AAAI (2022, 2024, 2025), NAACL (2024), COLM (2024), ICLR (2024), ACL (2023), KDD (2023), NLPCC (2024)

Reviewer: IEEE Transactions on Pattern Analysis and Machine Intelligence, Journal of Artificial Intelligence Research, Bioinformatics.

TEACHING

2024 Summer	TA: BSE740 Artificial Intelligence in Medicine, UIUC
2022 Spring	TA: CS598 Deep Learning for Healthcare, UIUC
2020 Summer	TA: Optimization Models and Applications, Tsinghua University
2020 Spring	TA: Bayesian Learning and Data Analysis, Tsinghua University
2019 Fall	TA: Learning from Data, Tsinghua University

(Updated on July 17, 2024.)