

Zifeng Wang

Co-Founder
Keiji AI Inc.
Seattle, WA, USA

zifengwang.ai@gmail.com
<https://zifeng.wang>
Google scholar: [kMIWwTAAAAAJ](https://scholar.google.com/citations?user=kMIWwTAAAAAJ)
+1 217-305-0065

EDUCATION

Ph.D. Computer Science, University of Illinois Urbana-Champaign, Urbana, USA, 2021-2025
M.S. Data Science, Tsinghua University, Beijing, China, 2018-2021
B.E. Engineering, Tongji University, Shanghai, China, 2014-2018

EXPERIENCE

2025- [Keiji AI Inc.](#)
Co-Founder, Head of AI

HIGHLIGHTS

I work for trustworthy AI-driven clinical development and decision-making through:

Biomedical Evidence Synthesis: I address the fragmentation and poor traceability of biomedical evidence by developing AI methods that retrieve, extract, integrate, and reason over literature, clinical trials, real-world data, and biomedical knowledge bases, enabling scalable evidence synthesis for drug discovery and translational medicine [[J3](#), [J5](#), [J7](#), [A13](#)].

Virtual Patients: I address patient heterogeneity and uncertain trial outcomes by developing virtual patient and trial models, enabling treatment response prediction, patient digital twins, and trial optimization [[A4](#), [A5](#), [A7](#), [A9](#), [A11](#), [A12](#)].

Clinical Foundation Models and Agents: I address the complexity and regulatory constraints of clinical workflows by developing foundation models and AI agents that execute trial design, patient recruitment, document generation, and clinical data science tasks with interpretable and verifiable outputs [[J1](#), [J4](#), [J6](#), [J8](#), [A6](#), [A10](#)].

SELECTED PAPERS

Journal Papers

- J8 **Wang, Z.**, Lin, J., Jin, Q., Gao, J., Pradeepkumar, J., Jiang, P., Lu, Z., Sun, J., "Benchmarking and Developing Large Language Models Using One Million Clinical Trials." In: *npj Digital Medicine* (2026)
- J7 **Wang, Z.**, Chen, Z., Yang, Z., Wang, X., Jin, Q., Peng, Y., Lu, Z., Sun, J., "Empowering Biomedical Evidence Exploration and Synthesis with Deep Knowledge Graph Research." In: *Nature Machine Intelligence* (2026)

- J6 **Wang, Z.**, Danek, B., Yang, Z., Chen, Z., Sun, J., "Making Large Language Models Reliable Data Science Programming Copilots for Biomedical Research." In: *Nature Biomedical Engineering* (2026)
- J5 **Wang, Z.**, Cao, L., Jin, Q., Chan, J., Wan, N., Afzali, B., Cho, H.-J., Choi, C.-I., Emamverdi, M., Gill, M. K., Kim, S.-H., Li, Y., Liu, Y., Luo, Y., Ong, H., Rousseau, J. F., Sheikh, I., Wei, J. J., Xu, Z., Zallek, C. M., Kim, K., Peng, Y., Lu, Z., Sun, J., "A foundation model for human-AI collaboration in medical literature mining." In: *Nature Communications* (2025)
- J4 **Wang, Z.**, Gao, J., Danek, B., Theodorou, B., Shaik, R., Thati, S., Won, S., Sun, J., "Compliance and Factuality of Large Language Models for Clinical Research Document Generation." In: *Journal of the American Medical Informatics Association* (2025)
- J3 **Wang, Z.**, Cao, L., Danek, B., Jin, Q., Lu, Z., Sun, J., "Accelerating Clinical Evidence Synthesis with Large Language Models." In: *npj Digital Medicine* (2025)
- J2 **Wang, Z.**, Wang, H., Danek, B., Li, Y., Mack, C., Arbuckle, L., Biswal, D., Poon, H., Wang, Y., Rajpurkar, P., Xiao, C., Sun, J., "A perspective for adapting generalist ai to specialized medical ai applications and their challenges." In: *npj Digital Medicine* (2025)
- J1 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: *Nature Communications* (2024)

Conference Papers

- A14 **Wang, Z.**, Wang, Z., Srinivasan, B., Ioannidis, V. N., Rangwala, H., Anubhai, R., "BioBridge: Bridging Biomedical Foundation Models via Knowledge Graphs." In: *ICLR'24*. 2024
- A13 Wen, Y., **Wang, Z.**, Sun, J., "Mindmap: Knowledge graph prompting sparks graph of thoughts in large language models." In: *ACL'24*. 2024
- A12 **Wang, Z.**, Gao, C., Xiao, C., Sun, J., "MediTab: Scaling Medical Tabular Data Predictors via Data Consolidation, Enrichment, and Refinement." In: *IJCAI'24*. 2024
- A11 Das, T., **Wang, Z.**, Sun, J., "Twin: Personalized clinical trial digital twin generation." In: *KDD'23*. 2023
- A10 **Wang, Z.**, Xiao, C., Sun, J., "AutoTrial: Prompting Language Models for Clinical Trial Design." In: *EMNLP'23*. 2023
- A9 **Wang, Z.**, Sun, J., "Transtab: Learning transferable tabular transformers across tables." In: *NeurIPS'22*. 2022
- A8 **Wang, Z.**, Wu, Z., Agarwal, D., Sun, J., "MedCLIP: Contrastive Learning from Unpaired Medical Images and Text." In: *EMNLP'22*. 2022
- A7 **Wang, Z.**, Sun, J., "PromptEHR: Conditional Electronic Healthcare Records Generation with Prompt Learning." In: *EMNLP'22*. 2022
- A6 **Wang, Z.**, Sun, J., "Trial2Vec: Zero-Shot Clinical Trial Document Similarity Search using Self-Supervision." In: *Findings of EMNLP'22*. 2022
- A5 **Wang, Z.**, Yang, Y., Wen, R., Chen, X., Huang, S.-L., Zheng, Y., "Lifelong learning based disease diagnosis on clinical notes." In: *PAKDD'21*. 2021
- A4 **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
- A3 **Wang, Z.**, Huang, S. L., Kuruoglu, E. E., Sun, J., Chen, X., Zheng, Y., "PAC-Bayes

- Information Bottleneck.” In: *ICLR’22*. 2022
- A2 **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*. 2020
- A1 **Wang, Z.**, Zhu, H., Dong, Z., He, X., Huang, S.-L., “Less is better: Unweighted data subsampling via influence function.” In: *AAAI’20*. 2020

AWARDS

- 2023 [Yee Memorial Fellowship](#), UIUC
- 2022 NeurIPS Scholar Award, NeurIPS
- 2022 [Yunni & Maxine Pao Memorial Fellowship](#), UIUC
- 2021 [Best Student Paper Award](#), PAKDD [A5]

MENTORING

UIUC

- 2024- Jiacheng Lin, PhD@UIUC, [Preprint’24](#), [Preprint’25](#)
- 2024- Pengcheng Jiang, MS@UIUC → PhD@UIUC, [COLM’25](#), [EMNLP’25](#), [NAACL’24](#)
- 2023- Lang Cao, MS@UIUC → PhD@UIUC, [Nat. Commun.’25](#), [npj Digit. Med.’25](#), [NAACL’24](#)
- 2023 Trisha Das, PhD@UIUC, [KDD’23](#), [Preprint’24](#)
- 2023 Chufan Gao, PhD@UIUC, [Preprint’24](#)

Visiting students

- 2023 Yilin Wen, MS → Baidu Research, [ACL’24](#)
- 2023 Ruiyu Wang, BS@UToronto → PhD@UToronto, [Preprint’23](#)
- 2023 Tao Feng, MS@Tsinghua → PhD@UIUC, [Preprint’23](#)

ACADEMIC SERVICE

Area Chair: COLM (2026); Program Committee/Reviewer: NeurIPS (2022, 2023, 2024, 2025), EMNLP (2022, 2023), IJCAI (2022, 2023, 2024), AAI (2022, 2024, 2025), NAACL (2024), COLM (2024, 2025), ICLR (2024, 2025), ICML (2025), ACL (2023), KDD (2023), NLPCC (2024)

Reviewer: Nature Communications, npj Digital Medicine, IEEE Transactions on Pattern Analysis and Machine Intelligence, Journal of Artificial Intelligence Research, Bioinformatics, Journal of the American Medical Informatics Association, ACM Computing Surveys

Co-host of [AI for Health Webinar](#).

TEACHING

Teaching Assistant

- 2022 Spring CS598 “Deep Learning for Healthcare”, UIUC

2020 Summer Graduate course: "Optimization Models and Applications", Tsinghua University
2020 Spring Graduate course: "Bayesian Learning and Data Analysis", Tsinghua University
2019 Fall Graduate course: "Learning from Data", Tsinghua University

INVITED TALKS

2026 "Empowering Clinical Development with LLMs and AI Agents", Invited talks: [DBMI@Columbia University](#), [BIDS@Yale University](#), [CSE@CUHK](#), [Tsinghua Medicine](#), [Duke-NUS Medicine](#), [CS@Zhejiang University](#), [CS@Tsinghua](#), [SIGS@Tsinghua](#), [AI4S@Peking University](#), [XtalPi Holdings](#)

2025 "Empowering Biomedicine with LLMs and AI Agents", Invited talks: [CS@Nanjing University](#), [AI@SJTU](#), [Shanghai AI Lab](#)

2025 "LLMs for hypothesis generation and validation in clinical trials", invited talks: [Joint Statistical Meetings](#), [AI for Health Webinar](#).

2025 "LLMs for clinical trial participant recruitment", [IT Roundtable@Clinical Research Forum](#).

2024 "Bridging biomedical foundation models", invited by [Xuegong Lab@Tsinghua University](#), [AI4Science@ByteDance](#).

2023 "Automate clinical trial design with large language models", invited by [Medidata](#).

2023 "Medical vision-language modeling", invited by [MedAI Group@Stanford University](#), [NCBI@NIH](#).

2023 "Unifying large language models and knowledge graphs", invited by [Bedrock@AWS](#).

SELECTED PRESS COVERAGE

2025 "Researchers have built a clinical trial database containing 1.65 million records, providing a data source for the development of vertical agents.", by [MIT Technology Review \(CN\)](#).

2025 "Clinical trials gain intelligence", by [Nature Biotechnology](#).

2025 "Health Science Top 25 of 2024", by [Nature Communications](#).

2024 "Using AI to match patients to clinical trials", by David LIM and Lauren GARDNER, [POLITICO](#).

2024 "How AI is being used to accelerate clinical trials", by Matthew Hutson, [Nature](#).

2024 "ChatGPT-Like AI Tool Promises to Speed Treatment Advances and Free Doctors' Time by Matching Patients with Clinical Trials ", by [Monica Bertagnolli](#), [the 17th Director of NIH](#).

2024 "NIH-developed AI algorithm matches potential volunteers to clinical trials", by [NIH News](#).

OTHER EXPERIENCE

2023 Amazon Web Service, Applied Scientist Intern, Mentor: [Zichen Wang](#)

2022 Amplitude, Research Intern, Mentor: [Cao Xiao](#)

2019-2021 Tencent, Research Intern, Mentor: [Yefeng Zheng](#)

(Updated on June 17, 2026.)