Qiao Jin, M.D.
qiao.jin@nih.gov
www.qiaojin.info
Research Fellow
National Library of Medicine
National Institutes of Health

I am a researcher working on AI for evidence-based medicine at NIH. My long-term goal is to democratize biomedical knowledge by providing accurate, verifiable, and understandable information to physicians, patients, and scientists. Currently, I work on language model evaluation, retrieval-augmented generation, language agents, and AI for clinical trials.

Education

2014–2022 **B.Sc.** (2019), **M.D.** (2022), Tsinghua University
Thesis: *Large-scale Text Mining from Biomedical Literature with Deep Neural Networks*
Supervisors: Dr. Sheng Yu (academic) & Dr. Jiahong Dong (clinical)
Overall GPA Ranking: 2/32

Research Experience

2025– **Research Fellow** (federal employee)
2022–2025 **Visiting Fellow** (intramural trainee)
National Library of Medicine, National Institutes of Health
Supervisor: Dr. Zhiyong Lu

2017–2019 **Visiting Research Scholar**
Department of Biomedical Informatics, University of Pittsburgh
Advisors: Dr. Xinghua Lu & Dr. William W. Cohen (CMU)

Awards & Honors

2024	Distinguished Poster Award	AMIA Annual Symposium
2024	Director's Challenge Innovation Award	National Institutes of Health
2023–2024	Fellows' Awards for Research Excellence	National Institutes of Health
2022	Outstanding Graduate Thesis (1/32)	Tsinghua University
2021	Best Clinical NLP Paper	IMIA Yearbook
2020–2021	Top Performance	TREC Biomedical Tracks
2019	First Place Winner	BioBank Disease AI Challenge
2015	National Scholarship	Ministry of Education of China
2015	Gold Medal	iGEM Competition
2013	Gold Medal	China Chemistry Olympiad

Selected Publications

All publications available at Google Scholar
† → Equal contribution

Language Model Evaluation

E1. **Jin, Qiao**, Chen, F., Zhou, Y., Xu, Z., Cheung, J. M., Chen, R., Summers, R. M., Rousseau, J. F., Ni, P., Landsman, M. J., Baxter, S. L., Al'Aref, S. J., Li, Y., Chen, A., Brejt, J. A., Chiang, M. F., Peng, Y. & Lu, Z. Hidden Flaws Behind Expert-Level Accuracy of Multimodal GPT-4 Vision in Medicine. *npj Digital Medicine*. https://www.nature.com/articles/s41746-024-01185-7 (2024).

E2. **Jin, Qiao**, Dhingra, B., Cohen, W. W. & Lu, X. Probing Biomedical Embeddings from Language Models. *Proceedings of the 3rd Workshop on Evaluating Vector Space Representations for NLP (RepEval)*. https://aclanthology.org/W19-2011 (2019).
– Pre-trained BioELMo, the first decoder-only language model in biomedicine.

E3. **Jin, Qiao**, Dhingra, B., Liu, Z., Cohen, W. W. & Lu, X. PubMedQA: A Dataset for Biomedical Research Question Answering. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. https://aclanthology.org/D19-1259 (2019).
– One of the most commonly used benchmarks for evaluating biomedical LLMs.
– Adopted by Google, Microsoft & OpenAI, Meta, and Anthropic.

AI Agents and Tool Learning

A1. Wang†, Z., **Jin†, Qiao**, Wei, C.-H., Tian, S., Lai, P.-T., Zhu, Q., Day, C.-P., Ross, C. & Lu, Z. GeneAgent: Self-verification Language Agent for Gene Set Knowledge Discovery using Domain Databases. *Nature Methods (to appear)* (2025).

A2. **Jin, Qiao**, Yang, Y., Chen, Q. & Lu, Z. GeneGPT: Augmenting Large Language Models with Domain Tools for Improved Access to Biomedical Information. *Bioinformatics* (2024).

A3. Khandekar†, N., **Jin†, Qiao**, Xiong†, G., Dunn, S., Applebaum, S. S., Anwar, Z., Sarfo-Gyamfi, M., Safranek, C. W., Anwar, A. A., Zhang, A., Gilson, A., Singer, M. B., Dave, A., Taylor, A., Zhang, A., Chen, Q. & Lu, Z. MedCalc-Bench: Evaluating Large Language Models for Medical Calculations. *NeurIPS Datasets and Benchmarks Track*. https://neurips.cc/virtual/2024/poster/97666 (2024).

Retrieval-Augmented Generation

R1. Xiong†, G., **Jin†, Qiao**, Lu, Z. & Zhang, A. Benchmarking Retrieval-Augmented Generation for Medicine. *Findings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*. https://aclanthology.org/2024.findings-acl.372/ (2024).
– MedRAG is a widely used toolkit for performing RAG in medicine.

R2. **Jin, Qiao**, Kim, W., Chen, Q., Comeau, D. C., Yeganova, L., Wilbur, J. & Lu, Z. MedCPT: Contrastive Pre-trained Transformers with Large-scale PubMed Search Logs for Zero-shot Biomedical Information Retrieval. *Bioinformatics* (2023).
– MedCPT has been downloaded over 2 million times on Hugging Face.

R3. Zhao†, Z., **Jin†, Qiao**, Chen, F., Peng, T. & Yu, S. A Large-scale Dataset of Patient Summaries for Retrieval-based Clinical Decision Support Systems. *Scientific Data*. https://www.nature.com/articles/s41597-023-02814-8 (2023).

AI for Clinical Trials

C1. **Jin, Qiao**, Wang, Z., Floudas, C., Chen, F., Gong, C., Bracken-Clarke, D., Xue, E., Yang, Y., Sun, J. & Lu, Z. Matching Patients to Clinical Trials with Large Language Models. *Nature Communications*. https://www.nature.com/articles/s41467-024-53081-z (2024).
– TrialGPT is the first framework that utilizes LLMs for clinical trial matching.
– Supported by the NIH Director's Challenge Innovation Award.
– Selected by the AI and Machine Learning focus of *Nature Communications*.
– Featured in the Health Science Top 25 of 2024 in *Nature Communications*.

C2. **Jin, Qiao**, Tan, C., Chen, M., Liu, X. & Huang, S. Predicting Clinical Trial Results by Implicit Evidence Integration. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. https://aclanthology.org/2020.emnlp-main.114/ (2020).
– Selected as the Best Clinical NLP Paper in 2020 by IMIA Yearbook.

Media Coverage

2024	POLITICO	Using AI to match patients to clinical trials. [C1]
2024	NIH Directors' Blog	ChatGPT-Like AI Tool Promises to Speed Treatment Advances and Free Doctors' Time by Matching Patients with Clinical Trials [C1]
2024	NIH News	NIH-developed AI algorithm matches potential volunteers to clinical trials [C1]
2024	AUA News	Connecting Patients to Clinical Trials With Artificial Intelligence [C1]
2024	MedScape	AI's Limitations in Clinical Decision-Making [E1]
2024	NIH News	NIH findings shed light on risks and benefits of integrating AI into medical decision-making. [E1]
2024	COSMOS	As we grapple with sovereign AI, perhaps we should treat computational resources as finite and precious [A2]

Academic Service

2025–	Area Chair	ACL Rolling Review (ARR)
2024–	Associate Editor	Journal of Medical Internet Research (JMIR)
2024–	Editorial Committee	Journal of Biomedical Informatics (JBI)
2024	Editorial Committee	Journal of the American Medical Informatics Association (JAMIA) Special Issue

Reviewer for Nature Medicine, Nature Methods, Nature Biomedical Engineering, Nature Computational Science, Nature Communications, npj Digital Medicine, NeurIPS, ICLR, ICML, SIGIR, ACM MM, ISMB, etc.

Last updated: April 3, 2025