Qiao Jin, M.D.

qiao.jin@nih.gov
https://qiaojin.info
Research Fellow
National Library of Medicine
National Institutes of Health

Education

2014–2022 **B.Sc.** (2019), **M.D.** (2022), Tsinghua University
Thesis: *Large-scale Text Mining from Biomedical Literature with Deep Neural Networks*
Mentors: 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
Mentor: Dr. Zhiyong Lu

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

Grant Funding

2025–2030 **Pathway to Independence (K99/R00) Award** National Institutes of Health
Total award amount: $997,000 K99LM014903
Improving Medical Information Verifiability with Generative Artificial Intelligence

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
2018	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

Publications

Google Scholar: 8,000+ citations
(65%+ from first/co-first-authored papers)

† Equal contribution
[n] Selected papers

[1] **Jin Q**, Wan N, Leaman R, Tian S, Wang Z, Yang Y, Wang Z, Xiong G, Lai P.-T, Zhu Q, Hou B, Sarfo-Gyamfi M, Zhang G, Gilson A, Bhasuran B, He Z, Zhang A, Sun J, Weng C, Summers RM, Chen Q, Peng Y & Lu Z. Entry-level Guide to the Use of Large Language Models for Medical Research. *Nature Protocols (in press)* (2026).

2 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. *Nature Machine Intelligence*. https://www.nature.com/articles/s42256-026-01266-0 (2026).

3 Bhasuran B, **Jin Q**, Deville A, Wu Y, Hanna K, Lu Z & He Z. A Curated Dataset for Question Answering on Laboratory Test Reference Ranges and Interpretation. *Scientific Data*. https://www.nature.com/articles/s41597-026-07554-z (2026).

4 Fang Y†, **Jin Q**†, Xiong G, Jin B, Zhong X, Ouyang S, Yang Y, Zhang A, Han J & Lu Z. Cell-o1: Training LLMs to Solve Single-Cell Reasoning Puzzles with Reinforcement Learning. *Bioinformatics*. https://pubmed.ncbi.nlm.nih.gov/42082388/ (2026).

5 Chan J, **Jin Q**, Wan N, Floudas CS, Xue E & Lu Z. Recommending Clinical Trials for Online Patient Cases Using Artificial Intelligence. *AMIA Annual Symposium Proceedings*. https://pubmed.ncbi.nlm.nih.gov/41726429/ (2025).

6 Wan NC, **Jin Q**, Chan J, Xiong G, Applebaum S, Gilson A, McMurry R, Taylor RA, Zhang A, Chen Q & Lu Z. Humans and Large Language Models in Clinical Decision Support: A Study with Medical Calculators. *AMIA Annual Symposium Proceedings*. https://pubmed.ncbi.nlm.nih.gov/41726433/ (2025).

[7] **Jin Q**, Wang Z, Yang Y, Zhu Q, Wright D, Huang T, Khandekar N, Wan N, Ai X, Wilbur WJ, He Z, Taylor RA, Chen Q & Lu Z. AgentMD: Empowering Language Agents for Risk Prediction with Large-Scale Clinical Tool Learning. *Nature Communications*. https://www.nature.com/articles/s41467-025-64430-x (2025).

8 Yang Y, **Jin Q**, Huang F & Lu Z. Adversarial Prompt and Fine-Tuning Attacks Threaten Medical Large Language Models. *Nature Communications*. https://www.nature.com/articles/s41467-025-64062-1 (2025).

9 Zhong X, Jin B, Ouyang S, Shen Y, **Jin Q**, Fang Y, Lu Z & Han J. *Benchmarking Retrieval-Augmented Generation for Chemistry* in *Second Conference on Language Modeling (COLM)* (2025). https://openreview.net/forum?id=qG4dL0bart.

10 Tang X, **Jin Q**, Zhu K, Yuan T, Zhang Y, Zhou W, Qu M, Zhao Y, Tang J, Zhang Z, Cohan A, Greenbaum D, Lu Z & Gerstein M. Risks of AI Scientists: Prioritizing Safeguarding over Autonomy. *Nature Communications*. https://www.nature.com/articles/s41467-025-63913-1 (2025).

11 Wang Z, Cao L, **Jin Q**, Chan J, Wan N, Afzali B, Cho H.-J, Choi C.-I, Emamverdi M, Gill MK, Kim S.-H, Li Y, Liu Y, Luo Y, Ong H, Rousseau JF, Sheikh I, Wei JJ, Xu Z, Zallek CM, Kim K, Peng Y, Lu Z & Sun J. A Foundation Model for Human-AI Collaboration in Medical Literature Mining. *Nature Communications*. https://www.nature.com/articles/s41467-025-62058-5 (2025).

12 Wang Z, Cao L, Danek B, **Jin Q**, Lu Z & Sun J. Accelerating Clinical Evidence Synthesis with Large Language Models. *npj Digital Medicine*. https://www.nature.com/articles/s41746-025-01840-7 (2025).

13 Yang Y, **Jin Q**, Zhu Q, Wang Z, Erramuspe Álvarez F, Wan N, Hou B & Lu Z. Beyond Multiple-Choice Accuracy: Real-World Challenges of Implementing Large Language Models in Healthcare. *Annual Review of Biomedical Data Science*. https://pubmed.ncbi.nlm.nih.gov/41031065/ (2025).

14 Zhu Q, Hou B, Mathai TS, Mukherjee P, **Jin Q**, Chen X, Wang Z, Cheng R, Summers RM & Lu Z. How Well Do Multimodal LLMs Interpret CT Scans? An Auto-Evaluation Framework for Analyses. *Journal of Biomedical Informatics*. https://pubmed.ncbi.nlm.nih.gov/40578543/ (2025).

15 Wang X, Tan M, **Jin Q**, Xiong G, Hu Y, Zhang A, Lu Z & Zhang M. MedCite: Can Language Models Generate Verifiable Text for Medicine? *Findings of the Association for Computational Linguistics: ACL 2025*. https://aclanthology.org/2025.findings-acl.967/ (2025).

[**16**] Wang Z†, **Jin Q**†, 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 Analysis Using Domain Databases. *Nature Methods*. https://www.nature.com/articles/s41592-025-02748-6 (2025).
– Featured in the Method to Watch 2025 by *Nature Methods*.
– Covered by NIH News and NIH Catalyst.

17 Zhang G, Xu Z, **Jin Q**, Chen F, Fang Y, Liu Y, Rousseau JF, Xu Z, Lu Z, Weng C & Peng Y. Leveraging Long Context in Retrieval Augmented Language Models for Medical Question Answering. *npj Digital Medicine*. https://www.nature.com/articles/s41746-025-01651-w (2025).

18 Chen Q, Hu Y, Peng X, Xie Q, **Jin Q**, Gilson A, Singer MB, Ai X, Lai P.-T, Wang Z, Keloth VK, Raja K, Huang J, He H, Lin F, Du J, Zhang R, Zheng WJ, Adelman RA, Lu Z & Xu H. Benchmarking Large Language Models for Biomedical Natural Language Processing Applications and Recommendations. *Nature Communications*. https://www.nature.com/articles/s41467-025-56989-2 (2025).

19 Bhasuran B, **Jin Q**, Xie Y, Yang C, Hanna K, Costa J, Shavor C, Han W, Lu Z & He Z. Preliminary Analysis of the Impact of Lab Results on Large Language Model Generated Differential Diagnoses. *npj Digital Medicine*. https://www.nature.com/articles/s41746-025-01556-8 (2025).

20 Xiong G†, **Jin Q**†, Wang X, Zhang M, Lu Z & Zhang A. Improving Retrieval-Augmented Generation in Medicine with Iterative Follow-up Questions. *Pacific Symposium on Biocomputing*. https://pubmed.ncbi.nlm.nih.gov/39670371/ (2025).

[**21**] Khandekar N†, **Jin Q**†, Xiong G†, Dunn S, Applebaum SS, Anwar Z, Sarfo-Gyamfi M, Safranek CW, Anwar AA, Zhang A, Gilson A, Singer MB, 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).
– Adopted by Anthropic, Nvidia, Sophont, Medical Sphere, and MedHELM.

[**22**] **Jin Q**, 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).

– 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*.
– Covered by NIH Catalyst, POLITICO, NLM Blog, NIH News, and Nature Index.

23 Zhang G, **Jin Q**, Zhou Y, Wang S, Idnay B, Luo Y, Park E, Nestor JG, Spotnitz ME, Soroush A, Campion TRJ, Lu Z, Weng C & Peng Y. Closing the Gap Between Open Source and Commercial Large Language Models for Medical Evidence Summarization. *npj Digital Medicine*. https://www.nature.com/articles/s41746-024-01239-w (2024).

[24] Xiong G†, **Jin Q**†, 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).

[25] **Jin Q**, Chen F, Zhou Y, Xu Z, Cheung JM, Chen R, Summers RM, Rousseau JF, Ni P, Landsman MJ, Baxter SL, Al'Aref SJ, Li Y, Chen A, Brejt JA, Chiang MF, 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).
– Covered by Medscape and NIH News.

26 Wei C-H†, Allot A†, Lai P.-T, Leaman R, Tian S, Luo L, **Jin Q**, Wang Z, Chen Q & Lu Z. PubTator 3.0: An AI-Powered Literature Resource for Unlocking Biomedical Knowledge. *Nucleic Acids Research*. https://pubmed.ncbi.nlm.nih.gov/38572754/ (2024).

27 Zhu Q, Chen X, **Jin Q**, Hou B, Mathai TS, Mukherjee P, Gao X, Summers RM & Lu Z. Leveraging Professional Radiologists' Expertise to Enhance LLMs' Evaluation for AI-Generated Radiology Reports. *2024 IEEE 12th International Conference on Healthcare Informatics (ICHI)*, 402–411. https://pubmed.ncbi.nlm.nih.gov/39691157/ (2024).

28 Zhang G, **Jin Q**, McInerney DJ, Chen Y, Wang F, Cole CL, Yang Q, Wang Y, Malin BA, Peleg M, Wallace BC, Lu Z, Weng C & Peng Y. Leveraging Generative AI for Clinical Evidence Synthesis Needs to Ensure Trustworthiness. *Journal of Biomedical Informatics*. https://pubmed.ncbi.nlm.nih.gov/38608915/ (2024).

29 He Z, Bhasuran B, **Jin Q**, Tian S, Hanna K, Shavor C, Garcia Arguello L, Murray P & Lu Z. Quality of Answers of Generative Large Language Models Versus Peer Users for Interpreting Laboratory Test Results for Lay Patients: Evaluation Study. *Journal of Medical Internet Research*. https://pubmed.ncbi.nlm.nih.gov/38630520/ (2024).

30 Keloth VK, Hu Y, Xie Q, Peng X, Wang Y, Zheng A, Selek M, Raja K, Wei C.-H, **Jin Q**, Lu Z, Chen Q & Xu H. Advancing Entity Recognition in Biomedicine via Instruction Tuning of Large Language Models. *Bioinformatics*. https://pubmed.ncbi.nlm.nih.gov/38514400/ (2024).

[31] **Jin Q**, Leaman R & Lu Z. PubMed and Beyond: Biomedical Literature Search in the Age of Artificial Intelligence. *eBioMedicine*. https://pubmed.ncbi.nlm.nih.gov/38306900/ (2024).

[32] **Jin Q**, Yang Y, Chen Q & Lu Z. GeneGPT: Augmenting Large Language Models with Domain Tools for Improved Access to Biomedical Information. *Bioinformatics*. https://pubmed.ncbi.nlm.nih.gov/38341654/ (2024).

33 Tian S, **Jin Q**, Yeganova L, Lai P.-T, Zhu Q, Chen X, Yang Y, Chen Q, Kim W, Comeau DC, Islamaj R, Kapoor A, Gao X & Lu Z. Opportunities and Challenges for ChatGPT and Large

Language Models in Biomedicine and Health. *Briefings in Bioinformatics*. https://pubmed.ncbi.nlm.nih.gov/38168838/ (2024).

[34] Zhao Z†, **Jin Q**†, 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).

[35] **Jin Q**, Kim W, Chen Q, Comeau DC, Yeganova L, Wilbur J & Lu Z. MedCPT: Contrastive Pre-Trained Transformers with Large-Scale PubMed Search Logs for Zero-Shot Biomedical Information Retrieval. *Bioinformatics*. https://pubmed.ncbi.nlm.nih.gov/37930897/ (2023).
– Downloaded over 8 million times on Hugging Face.
– Supporting NLM products that serve millions of user requests.

36 Yuan Z, **Jin Q**, Tan C, Zhao Z, Yuan H, Huang F & Huang S. RAMM: Retrieval-Augmented Biomedical Visual Question Answering with Multi-Modal Pre-Training. *Proceedings of the 31st ACM International Conference on Multimedia*. https://dl.acm.org/doi/10.1145/3581783.3611830 (2023).

37 Shin A, **Jin Q** & Lu Z. Multi-stage Literature Retrieval System Trained by PubMed Search Logs for Biomedical Question Answering. *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023)*. https://ceur-ws.org/Vol-3497/paper-016.pdf (2023).

[38] **Jin Q**, Leaman R & Lu Z. Retrieve, Summarize, and Verify: How Will ChatGPT Affect Information Seeking from the Medical Literature? *Journal of the American Society of Nephrology*. https://pubmed.ncbi.nlm.nih.gov/37254254/ (2023).

[39] **Jin Q**, Shin A & Lu Z. LADER: Log-Augmented DEnse Retrieval for Biomedical Literature Search. *The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. https://dl.acm.org/doi/10.1145/3539618.3592005 (2023).

[40] **Jin Q**, Yuan Z, Xiong G, Yu Q, Ying H, Tan C, Chen M, Huang S, Liu X & Yu S. Biomedical Question Answering: A Survey of Approaches and Challenges. *ACM Computing Surveys*. https://dl.acm.org/doi/10.1145/3490238 (2022).

41 **Jin Q**, Tan C, Chen M, Yan M, Zhang N, Huang S & Liu X. State-of-the-Art Evidence Retriever for Precision Medicine: Algorithm Development and Validation. *JMIR Medical Informatics*. https://pubmed.ncbi.nlm.nih.gov/36409468/ (2022).

42 Li G, Zhou Y, Ji J, Liu X, **Jin Q** & Zhang L. Surging Publications on the COVID-19 Pandemic. *Clinical Microbiology and Infection*. https://pubmed.ncbi.nlm.nih.gov/32977004/ (2021).

[43] **Jin Q**, 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.

44 Zhou Y, Ren Q, Chen G, **Jin Q**, Cui Q, Luo H, Zheng K, Qin Y & Li X. Chronic Kidney Diseases and Acute Kidney Injury in Patients With COVID-19: Evidence From a Meta-Analysis. *Frontiers in Medicine*. https://pubmed.ncbi.nlm.nih.gov/33224965/ (2020).

45 **Jin Q**, Ding H, Li L, Huang H, Wang L & Yan J. Tackling MeSH Indexing Dataset Shift with Time-Aware Concept Embedding Learning. *Database Systems for Advanced Applica-*

tions (DASFAA). https://dl.acm.org/doi/abs/10.1007/978-3-030-59419-0_29 (2020).

[**46**] **Jin Q**, Dhingra B, Cohen WW & 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).

[**47**] **Jin Q**, Dhingra B, Liu Z, Cohen WW & 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).
– Adopted by Google, Microsoft & OpenAI, Meta, Anthropic, and MedHELM.

48 **Jin Q**, Liu J & Lu X. Deep Contextualized Biomedical Abbreviation Expansion. *Proceedings of the 18th BioNLP Workshop and Shared Task*. https://aclanthology.org/W19-5010/ (2019).

49 Wang H, Liu X, Tao Y, Ye W, **Jin Q**, Cohen WW & Xing EP. Automatic Human-like Mining and Constructing Reliable Genetic Association Database with Deep Reinforcement Learning. *Pacific Symposium on Biocomputing*. https://pubmed.ncbi.nlm.nih.gov/30864315/ (2019).

50 **Jin Q**, Dhingra B, Cohen WW & Lu X. AttentionMeSH: Simple, Effective and Interpretable Automatic MeSH Indexer. *Proceedings of the 6th BioASQ Workshop*. https://aclanthology.org/W18-5306/ (2018).

51 Dhingra B, **Jin Q**, Yang Z, Cohen WW & Salakhutdinov R. Neural Models for Reasoning over Multiple Mentions Using Coreference. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): Human Language Technologies*. https://aclanthology.org/N18-2007/ (2018).

Invited Talks

2026	AIM-AHEAD AI Optimization Subcore	Artificial Intelligence for Evidence-Based Medicine
2026	University of Minnesota	Med-V1: Small Language Models for Zero-shot and Scalable Biomedical Evidence Attribution
2025	Yale University Rising Star Seminar	Language Modeling for Biomedicine: From PubMedQA to TrialGPT
2025	NCI Office of Data Sharing	Artificial Intelligence for Evidence-Based Medicine
2025	Pitt/UPMC GenAI and LLMs Interest Group	Artificial Intelligence for Evidence-Based Medicine
2025	NIH AI Club	Artificial Intelligence for Evidence-Based Medicine
2025	Pacific Symposium on Biocomputing	Artificial Intelligence for Evidence-Based Medicine
2024	Weill Cornell Medicine	Language Modeling for Biomedicine: Evaluation, Augmentation, and Application
2024	OHDSI NLP Work Group	Language Modeling for Biomedicine: Evaluation, Augmentation, and Application

2024	ISMB Tutorial	A Practical Introduction to Large Language Models in Biomedical Data Science Research
2024	Tsinghua University	Language Modeling for Biomedicine: Evaluation, Augmentation, and Application
2024	NLM Town Hall	Generative AI Panel Discussion
2024	Ci4CC Community Call	Language Modeling for Biomedicine: Evaluation, Augmentation, and Application
2023	University of Delaware	Applying Large Language Models to Biomedicine
2023	VA AI Health Summit	Applying Large Language Models to Biomedicine
2023	NIH Special Interest Group for NLP	Overview of ChatGPT and other Large Language Models and their applications in Biomedicine

Academic Service

2026–	Area Chair	Conference on Neural Information Processing Systems (NeurIPS)
2026–	Editor	Journal of the American Medical Informatics Association (JAMIA)
2025–	Area Chair	ACL Rolling Review (ARR)
2024–	Associate Editor	Journal of Medical Internet Research (JMIR)
2024–25	Student Editor	Journal of Biomedical Informatics (JBI)
2024	Editor	JAMIA Special Issue on ChatGPT and Large Language Models (LLMs) in Biomedicine and Health

Ad-hoc reviewer for Nature Medicine, Nature Methods, Nature Biomedical Engineering, Nature Computational Science, Nature Communications, Nature Reviews Bioengineering, npj Digital Medicine, npj Health Systems, The Lancet Regional Health, The Lancet Digital Health, Cell Reports Medicine, Patterns, NEJM AI, ICLR, ICML, CVPR, SIGIR, ACM Multimedia, ISMB, etc.

Last updated: July 22, 2026