

YANJUN GAO, PhD

Assistant Professor
Department of Biomedical Informatics
University of Colorado, Anschutz Medical Campus

personal website: serenayj.github.io
lab website: www.larknlp.com
yanjun.gao@cuanschutz.edu

EDUCATION

Ph.D. Computer Science and Engineering
Pennsylvania State University, State College, PA, USA 2014 – 2021
Advisor: Dr. Rebecca J. Passonneau
Field of Study: Natural Language Processing (NLP)
THESIS - *Analysis of Text to Identify, Represent and Group Distinct Propositions*

Bachelor of Management, Electronic Commerce
Harbin Finance University, Harbin, China 2010 – 2014
GPA Ranking: Top 1%
China National Scholarship

Minor, English
Heilongjiang University, Harbin, China 2010 – 2014
Second Foreign Language: Japanese

PROFESSIONAL EXPERIENCE

Tenure-Track Assistant Professor, Principal Investigator September 2024 – Present
Language, Reasoning, Knowledge (LARK) Lab,
Department of Biomedical Informatics,
University of Colorado Anschutz Medical Campus

- Founded the LARK Lab to conduct research for novel methods that leverage natural language processing (NLP) and machine learning to solve critical healthcare problems, including diagnostic decision making, patient education, and bias evaluation in clinical predictions.
- Established a multidisciplinary team focusing on developing and evaluating large language models (LLMs) for applications such as automated medical knowledge graph integration, uncertainty estimation, and explainable medical diagnostics systems.
- Secured external funding, including an NIH NLM K99/R00 award, to advance multimodal clinical NLP models for diagnosis generation and patient summarization.
- Developed collaborative partnerships with clinicians and computer scientists to design technical solutions that align with real-world healthcare needs.
- Served in the Department of Biomedical Informatics Seminar Committee to organize seminar talks.
- Mentored graduate students and research staff to cultivate skills in cutting-edge NLP techniques and foster innovative contributions to medical informatics.

Postdoctoral Research Associate August 2021 – August 2024
University of Wisconsin Madison, Madison, WI

- Authored a systematic review on clinical NLP tasks, developed a novel annotation framework for electronic health records and designed a new suite of tasks for clinical NLP.
- Pioneered the first clinical NLP benchmark for diagnostic reasoning, establishing baseline results using the leading NLP methods including fine-tuning LLMs, and innovated a novel knowledge graph prompting method for LLMs in diagnosis prediction, mentored PhD students to design human and automated evaluation metrics for diagnosis prediction.
- Designed an automated prompt engineering framework for healthcare applications and published papers detailing these advancements.

- Served as NLP consultant to the University of Wisconsin Hospital (UW Health) operational team to design human-factor prompting and evaluation framework for LLM healthcare use cases.

Research Assistant, Pennsylvania State University, State College, PA

NLP Lab, Department of Computer Science, 2017 – 2021

- Developed “PyrEval”, an innovative method and software package for summarization evaluation, achieving higher human correlation than conventional metrics such as ROUGE.
- Developed graph-based methods for discourse connective predictions and decomposition of complex sentences into individual propositions
- Collaborated with researchers in education and collected data from real-world classrooms on student writings, conducted reliability studies on automated pyramid summarization evaluation in student writing and authored papers to publish these findings.

School of Information Science and Technology 2015 – 2016

- Designed and implemented a multi-user system for collaborative information searching, sharing, and decision-making in an online environment. Conducted pilot studies, performed quantitative and qualitative analysis and published results.

Research Intern, Kwai Inc (Kuaishou Technology)

May 2019 – August 2019

Silicon Valley AI Lab, Palo Alto, CA

- Designed and implemented an advanced multi-modal pipeline to fuse video-understanding neural networks and text encoders, facilitating video-language understanding and interaction in temporal grounding tasks. Conducted comprehensive experiments and successfully published the results.

GRANTS

National Library of Medicine K99/R00 Pathway to Independence Award

2023 – 2027

National Institutes of Health, “Developing and Evaluating Multi-Modal Clinical Diagnostic Reasoning Models for Automated Diagnosis Generation”, Amount: \$ 874,800

CU Anschutz Faculty Wellbeing Innovation Award.

2025 – 2027

CU Anschutz, “Optimizing AI-Powered In-Basket Messaging to Improve Providers Well-Being”, Amount: \$ 50,000

PATENTS

System and Method for Generating Personalized Health Infographics.

Stonbraker SB, **Gao Y**, Cato K. U.S. Provisional Patent Application No. 63/853,186. Filed July 29, 2025. Applicant: The Regents of the University of Colorado, a body corporate.

Video processing method and electronic device.

Gao Y, Chen X, Wang T, International Patent Publication No. WO2022134634A1. Published June 30, 2022. Assignee: Beijing Dajia Internet Information Technology Co., Ltd.

JOURNAL PUBLICATIONS

- Yoon W, Chen S, **Gao Y**, Zhao Z, Dligach D, Bitterman DS, Afshar M, Miller T. LCD benchmark: long clinical document benchmark on mortality prediction for language models. *J Am Med Inform Assoc.* 2025;32(2):285–295. doi:10.1093/jamia/ocae287
- Eslami B, Afshar M, Tootooni S, Miller TA, Churpek MM, **Gao Y**, Dligach D. Toward digital twins in the intensive care unit: a medication management case study. *J Am Med Inform Assoc.* 2025;ocaf127. doi:10.1093/jamia/ocaf127
- Croxford E, **Gao Y**, First E, Pellegrino N, Schnier M, Caskey J, Oguss M, Wills G, Chen G, Dligach D, Churpek MM, Mayampurath A, Liao F, Goswami C, Wong KK, Patterson BW, Afshar M. Evaluating clinical AI summaries with large language models as judges. *NPJ Digit Med.* 2025;8(1):2025. doi:10.1101/2025.04.22.25326219

- Croxford E, **Gao Y**, Pellegrino N, Wong KK, Wills G, First E, Schnier M, Burton K, Ebby C, Gorski J, Kalscheur M, Khalil S, Pisani M, Rubeor T, Stetson P, Liao F, Goswami C, Patterson B, Afshar M. Development and validation of the provider documentation summarization quality instrument for large language models. *J Am Med Inform Assoc*. 2025;32(6):1050–1060. doi:10.1093/jamia/ocaf068
- Croxford E, **Gao Y**, Pellegrino N, Wong KK, Wills G, First E, Liao F, Goswami C, Patterson B, Afshar M. Current and future state of evaluation of large language models for medical summarization tasks. *NPJ Health Syst*. 2025;2(1):6. doi:10.1038/s44401-024-00011-2
- **Gao Y**, Myers S, Chen S, Dligach D, Miller T, Bitterman DS, Chen G, Mayampurath A, Churpek MM, Afshar M. Uncertainty estimation in diagnosis generation from large language models: next-word probability is not pre-test probability. *JAMIA Open*. 2025;8(1):ooae154. doi:10.1093/jamiaopen/ooae154
- Myers S, Miller TA, **Gao Y**, Churpek MM, Mayampurath A, Dligach D, Afshar M. Lessons learned on information retrieval in electronic health records: a comparison of embedding models and pooling strategies. *J Am Med Inform Assoc*. 2025;32(2):357–364. doi:10.1093/jamia/ocae308
- **Gao Y**, Li R, Croxford E, Tesch S, To D, Caskey J, Patterson BW, Churpek MM, Miller T, Dligach D, Afshar M. Leveraging a medical knowledge graph into large language models for diagnosis prediction: design and application study. *J Med Internet Res AI*. 2025; e58670. doi:10.2196/58670
- Afshar M, **Gao Y**, Gupta D, Croxford E, Demner-Fushman D. On the role of the UMLS in supporting diagnosis generation proposed by large language models. *J Biomed Inform*. 2024;157:104707. doi:10.1016/j.jbi.2024.104707
- Afshar M, **Gao Y**, Wills G, Wang J, Churpek MM, Westenberger CJ, Kunstman DT, Gordon JE, Liao FJ, Patterson BW. Prompt engineering GPT-4 to answer patient inquiries: a real-time implementation in the electronic health record across provider clinics. *JAMIA Open*. 2024;7(1):2024–01. doi:10.1093/jamiaopen/ooae080.
- **Gao Y**, Mahajan D, Uzuner Ö, Yetisgen M. Clinical natural language processing for secondary uses. *J Biomed Inform*. 2024;146:104596. doi:10.1016/j.jbi.2024.104596.
- **Gao Y**, Dligach D, Miller T, Churpek MM, Uzuner Ö, Afshar M. Progress note understanding—assessment and plan reasoning: overview of the 2022 N2C2 Track 3 shared task. *J Biomed Inform*. 2023 Jun;104346. doi:10.1016/j.jbi.2023.104346.
- **Gao Y**, Dligach D, Miller T, Caskey J, Sharma B, Churpek MM, Afshar M. DR.BENCH: diagnostic reasoning benchmark for clinical natural language processing. *J Biomed Inform*. 2023;138:104286. doi:10.1016/j.jbi.2023.104286
- **Gao Y**, Dligach D, Christensen L, Tesch S, Laffin R, Xu D, Miller T, Uzuner Ö, Churpek MM, Afshar M. A scoping review of publicly available language tasks in clinical natural language processing. *J Am Med Inform Assoc*. 2022 Sep 12;29(10):1797–1806. doi:10.1093/jamia/ocac127
- Yetisgen M, Uzuner Ö, **Gao Y**, Mahajan D. Call for papers: special issue on clinical natural language processing for secondary use applications. *J Biomed Inform*. 2022;133:104152. doi:10.1016/j.jbi.2022.104152
- Davies PM, Passonneau RJ, Muresan S, **Gao Y**. Analytical techniques for developing argumentative writing in STEM: a pilot study. *IEEE Trans Educ*. 2021;65(3):373–383.

CONFERENCE AND OTHER PUBLICATIONS (Selected)

*Corresponding author

- Kruse M, Afshar M, Khatwani S, Mayampurath A, Chen G, **Gao Y***. Simple yet effective: an information-theoretic approach to multi-LLM uncertainty quantification. In: *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP Main 2025)*. 2025.
- Kruse M, Hu S, Derby N, Wu Y, Stonbraker S, Yao B, Wang D, Goldberg E, **Gao Y***. Large language models with temporal reasoning for longitudinal clinical summarization and prediction. In: *Findings of the Association for Computational Linguistics: EMNLP Findings 2025*. 2025.
- Khatwani S, Cheng H, Afshar M, Dligach D, **Gao Y***. Brittleness and promise: knowledge graph-based reward modeling for diagnostic reasoning. In: *Proceedings of the GenAI4Health Workshop at the 38th Annual Conference on Neural Information Processing Systems (GenAI4Health @ NeurIPS 2025)*. 2025.
- Li R, Chen C, Hu Y, **Gao Y**, Wang X, Yilmaz E. Attributing response to context: a Jensen–Shannon divergence-driven mechanistic study of context attribution in retrieval-augmented generation. In: *Proceedings of the Conference on Language Modeling: The First Workshop on the Application of LLM Explainability to Reasoning and Planning (XLLM-Reason-Plan @ COLM 2025)*. 2025. **Best Paper**.
- Li R, **Gao Y***. Anchored answers: unravelling positional bias in GPT-2’s multiple-choice questions. In: *Findings of the Association for Computational Linguistics (ACL Findings 2025)*. 2025.

- **Gao Y***, Myers S, Chen S, Dligach D, Miller TA, Bitterman DS, Mayampurath A, Churpek MM, Afshar M. Position paper on diagnostic uncertainty estimation from large language models: next-word probability is not pre-test probability. In: *Proceedings of the GenAI4Health Workshop at NeurIPS 2024*. 2024.
- **Gao Y***, Myers S, Chen S, Dligach D, Miller TA, Bitterman DS, Churpek MM, Afshar M. When raw data prevails: are large language model embeddings effective in numerical data representation for medical machine learning applications? In: *Findings of Empirical Methods in Natural Language Processing (EMNLP Findings 2024)*. 2024.
- Chen, Xin, Hanxian Huang, **YanJun Gao**, Yi Wang, Jishen Zhao, and Ke Ding. Learning to Maximize Mutual Information for Chain-of-Thought Distillation. *Findings of Association of Computational Linguistics (ACL Findings 2024)*.
- Croxford E, **Gao Y**, Patterson B, To D, Tesch S, Dligach D, Mayampurath A, Churpek MM, Afshar M. Development of a human evaluation framework and correlation with automated metrics for natural language generation of medical diagnoses. In: *Proceedings of the American Medical Informatics Association Annual Symposium (AMIA 2024)*. 2024.
- **Gao Y**, Dligach D, Miller T, Churpek MM, Afshar M. Overview of the Problem List Summarization (ProbSum) 2023 Shared Task on summarizing patients' active diagnoses and problems from electronic health record progress notes. In: *Proceedings of the 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks (BioNLP 2023)*. 2023.
- Sharma B, **Gao Y**, Miller T, Churpek MM, Afshar M, Dligach D. Multi-task training with in-domain language models for diagnostic reasoning. In: *Proceedings of the 5th Clinical Natural Language Processing Workshop (ClinicalNLP 2023)*. 2023.
- **Gao Y**, Dligach D, Miller T, Churpek MM, Afshar M. Summarizing patients' problems from hospital progress notes using pre-trained sequence-to-sequence models. In: *Proceedings of the 29th International Conference on Computational Linguistics (COLING 2022)*. 2022.
- **Gao Y**, Dligach D, Miller T, Tesch S, Laffin R, Churpek MM, Afshar M. Hierarchical annotation for building a suite of clinical natural language processing tasks: progress note understanding. In: *Proceedings of the 13th Language Resources and Evaluation Conference (LREC 2022)*. 2022.
- **Gao, Yanjun**, Kenneth Ting-hao Huang, Rebecca J. Passonneau. "ABCD: A Graph Framework to Convert Complex Sentences to a Covering Set of Simple Sentences". *59th Annual Meeting of the Association for Computational Linguistics (ACL Main 2021)*.
- **Gao Y**, Sun C, Passonneau RJ. Automatic pyramid summarization evaluation. In: *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL 2019)*. 2019.

DISTINCTIONS

- Best Paper, The First Workshop on the Application of LLM Explainability to Reasoning and Planning at Conference on Language Modeling 2025.
- Biomedical Informatics Year-in-Review (nomination), American Medical Informatics Association (AMIA) Annual Symposium 2023.
- Outstanding Reviewer, npj Digital Medicine, Nature 2023.
- Trainee Outstanding Research Awards at Department of Medicine, University of Wisconsin Madison 2023.
- Best Clinical Research at Department of Medicine Annual Research Day, University of Wisconsin Madison 2022.
- Best Poster in New York Academy of Sciences NLP, Dialog and Speech Symposium (NYAS NDS) 2019
- National Scholarship, Ministry of Education in China 2013 – 2014.

EDITORIAL BOARD

Journal of Biomedical Informatics (starting 2026)

AREA CHAIR

Association of Computational Linguistics

ORGANIZING COMMITTEE

International Workshop on Graph-based Natural Language Processing (TextGraphs 2022, 2024)
 Biomedical Natural Language Processing Workshop (BioNLP 2023)
 National Natural Language Processing Clinical Challenge (N2C2 2022)

PROGRAM COMMITTEE

GenAI4Health Workshop 2025
International Conference on Computational Linguistics (COLING 2025)
International Conference on Artificial Intelligence in Education (AIED 2022)
International Conference on Language Resources and Evaluation (LREC 2022, 2024)
International Workshop on Evaluation and Comparison for NLP systems (Eval4NLP 2020-2022)
China National Conference on Computational Linguistics (CCL 2020)

REVIEWER

National Institutes of Health Center for Scientific Review (TOPICS IN CLINICAL DATA MANAGEMENT, ANALYSIS, INFORMATICS AND DIGITAL HEALTH A, ZRG1 HSS - R)
Association of Computational Linguistics (ACL)
International Conference Artificial Intelligence in Education (AIED)
International Conference on Language Resources and Evaluation (LREC)
Journal of Language Resources and Evaluation
Journal of Biomedical Informatics (JBI)
Journal of American Medical Informatics Association (JAMIA)
Journal of Healthcare Informatics Research
npj Digital Medicine, Nature
Scientific Reports, Nature
BMC medical informatics and decision making

TEACHING EXPERIENCES

Guest Lecturer, Department of Computer Science and Engineering, Pennsylvania State University
CSE 597, Natural Language Processing Spring 2018 – Spring 2021

Teaching Assistant, Department of Computer Science and Engineering, Pennsylvania State University
CMPSC/DS 442, Artificial Intelligence Fall 2018 – Spring 2021
CMPSC 221, Introduction to C++ Fall 2016 – Fall 2017

MENTORING EXPERIENCES

Postdoctoral Scholars

- He Cheng (Ph.D. in Computer Science), 2025 – Present
- Minh Son Nguyen (Ph.D. in Computer Science), 2025 – Present

Graduate Students

- Alexander Kotz, PhD student, Computational Bioscience Program, University of Colorado Anschutz, 2025 – Present (Role: PhD committee member)
- Parker Hicks, PhD student, Human Medical Genetics and Genomics Program, University of Colorado Anschutz, 2024 – Present (Role: PhD committee member)
- Emma Croxford, PhD student, Biomedical Data Science, University of Wisconsin-Madison, 2022 – Present (Role: PhD committee member, NLP methods advisor)
- Shiyue Hu, master student, Computational Linguistics, University of Colorado Boulder, 2024 – Present (Role: master thesis advisor)
- Saksham Khatwani, master student, Computer Science, University of Colorado Boulder, 2025 – Present (Role: master thesis advisor)
- Jason Wang, master student, Computer Science, University of California, Berkeley, 2021 – 2023 (Role: master thesis advisor)

Research Undergraduate Students (pre-2023): Mentored five undergraduate students and one honors group at Pennsylvania State University (2017–2022).