

Tian Kang

Updated May 10, 2021

Email: tiankang.nlp@gmail.com

Phone: (917) 288-9570

GitHub: [@Tian312](#)

Affiliation: Biomedical Informatics, Columbia University

Address: 622 West 168 Street, New York, NY

Website: <http://www.tiankangnlp.com>

Research summary

I'm a PhD student at the department of Biomedical Informatics at Columbia University with an interest in contributing Natural Language Processing methods to facilitate healthcare. I'm graduating with a PhD degree in June, 2021. In long term, my goal is to develop advanced natural language processing and natural language understanding technologies to enable a cognitive computing healthcare systems.

Education

Columbia University New York, NY
Ph.D. in Biomedical Informatics GPA: 3.90/4.00 2021
Ph.D. Thesis: *Towards Unified Medical Evidence Computation from Literature for Evidence-based Medicine*
Mentor: [Dr. Chunhua Weng](#)

Columbia University New York, NY
M.A. in Biomedical Informatics GPA: 3.76/4.00 2016
M.A. Thesis: *EliIE: A Machine Learning-based Information Extraction System to Formalize Clinical Research Eligibility Criteria into OMOP Common Data Model*

Huazhong University of Science and Technology Hubei, China
B.S. in Bioinformatics GPA: 3.50/4.00 2014

Publications

1. **Kang, T.**, Turfah, A. Kim, J. Perotte, A. and Weng, C. (2021). A Neuro-Symbolic Method for Understanding Free-text Medical Evidence. *Journal of the American Medical Informatics Association*, ocab077, <https://doi.org/10.1093/jamia/ocab077> (GitHub)
2. **Kang, T.**, Perotte, A., Tang, Y., Ta, C. and Weng, C. (2021). [UMLS-based Data Augmentation for Biomedical NLP with Limited Data](#). *Journal of the American Medical Informatics Association*, 28(4), 812-823. (GitHub)
3. **Kang, T.**, Zou, S., and Weng, C. (2019). [Pretraining to recognize PICO elements from randomized controlled trial literature](#). *Studies in health technology and informatics*, 264, 188. (GitHub)

4. **Kang, T.**, Zhang, S., Tang, Y., Hruby, G. W., Rusanov, A., Elhadad, N., and Weng, C. (2017). [EliIE: An open-source information extraction system for clinical trial eligibility criteria](#). *Journal of the American Medical Informatics Association*, 24(6), 1062-1071. (GitHub)
5. **Kang, T.**, Zhang, S., Xu, N., Wen, D., Zhang, X., and Lei, J. (2017). [Detecting negation and scope in Chinese clinical notes using character and word embedding](#). *Computer methods and programs in biomedicine*, 140, 53-59.
6. **Kang, T.**, Elhadad, N., and Weng, C. (2015). [Initial readability assessment of clinical trial eligibility criteria](#). In *AMIA Annual Symposium Proceedings* (Vol. 2015, p. 687). American Medical Informatics Association.
7. Wei, D. H., **Kang, T.**, Pincus, H. A., and Weng, C. (2019). [Construction of disease similarity networks using concept embedding and ontology](#). *Studies in health technology and informatics*, 264, 442.
8. Rogers, J. R., Callahan, T. J., **Kang, T.**, Bauck, A., Khare, R., Brown, J. S., ... and Weng, C. (2019). [A Data Element-Function Conceptual Model for Data Quality Checks](#). *eGEMs*, 7(1).
9. Yuan, C., Ryan, P.B., Ta, C., Guo, Y., Li, Z., Hardin, J., Makadia, R., Jin, P., Shang, N., **Kang, T.** and Weng, C., (2019). [Criteria2Query: a natural language interface to clinical databases for cohort definition](#). *Journal of the American Medical Informatics Association*, 26(4), 294-305.
10. Butler, A., Wei, W., Yuan, C., **Kang, T.**, Si, Y., and Weng, C. (2018). [The data gap in the EHR for clinical research eligibility screening](#). *AMIA Summits on Translational Science Proceedings*, 2018, 320.
11. Sen, A., Goldstein, A., Chakrabarti, S., Shang, N., **Kang, T.**, Yaman, A., Ryan, P.B. and Weng, C.(2018). [The representativeness of eligible patients in type 2 diabetes trials: a case study using GIST 2.0](#). *Journal of the American Medical Informatics Association*, 25(3), 239-247.
12. Zhang, S., **Kang, T.**, Qiu, L., Zhang, W., Yu, Y., and Elhadad, N. (2017, April). [Cataloging treatments discussed and used in online autism communities](#). In *Proceedings of the 26th International Conference on World Wide Web* (pp. 123-131).
13. Zhang, S., **Kang, T.**, Zhang, X., Wen, D., Elhadad, N., and Lei, J. (2016). [Speculation detection for Chinese clinical notes: impacts of word segmentation and embedding models](#). *Journal of biomedical informatics*, 60, 334-341.

14. Zhou, F., Cao, H., Zuo, X., Zhang, T., Zhang, X., Liu, X., Xu, R. Chen, G., Zhang, Y., Zheng, X., Jin, X., Gao, J., Mei, J., Sheng, Y., Li, Q., Liang, B., Shen, J., Shen, C., Jiang, H., Zhu, C., Fan, X., Xu, F., Yue, M., Yin, X., Ye, C., Zhang, C., Liu, X., Yu, L., Wu, J., Chen, M., Zhuang, X., Tang, L., Shao, H., Wu, L., Li, J., Xu, Y., Zhang, Y., Zhao, S., Wang, Y., Li, G., Xu, H., Zeng, L., Wang, J., Bai, M., Chen, Y., Chen, W., **Kang, T.**, Wu, Y., Xu, X., Zhu, Z., Cui, Y., Wang, Z., Yang, C., Wang, P., Xiang, L., Chen, X., Zhang, A., Gao, X., Zhang, F., Xu, J., Zheng, M., Zheng, J., Zhang, J., Yu, X., Li, Y., Yang, S., Yang, H., Wang, J., Liu, J., Hammarström, L., Sun, L., Wang, J. and Zhang, X. (2016). [Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease](#). *Nature genetics*, 48(7), 740-746.

Papers under review
and in progress

1. **Kang, T.**, Perotte, A., Tang, Y., and Weng, C. Integrating Sequence Labeling and Pattern Recognition to Simplify Composite Mentions.
2. **Kang, T.**, Perotte, A., Jaehyun, K., Turfah, A., and Weng, C. EvidenceMap: Representing Medical Evidence in Medical Literature to Advance Evidence Navigation and Synthesis.
3. **Kang, T.**, Perotte, A., Jaehyun, K., Rogers, J., and Weng, C. PubMed in EvidenceMap: A Multi-Level Indexed Medical Evidence Base for Assisting Evidence-based Medicine.

Talks and tutorials

Pretraining to recognize PICO elements from randomized controlled trial literature.

MEDINFO 2019, Lyon, France

August 2019

EliIE: An open-source information extraction system for clinical trial eligibility criteria.

JAMIA Journal Club Webinar

October 2017

Cataloguing treatments discussed and used in online autism communities.

International World Wide Web Conference, Perth, Australia

April 2017

"We make choices we think are going to save us": Debate and stance identification for online breast cancer CAM discussions.

International World Wide Web Conference, Perth, Australia

April 2017

Initial Readability Assessment of Clinical Trial Eligibility Criteria.

AMIA Annual Symposium, San Francisco, United States

November 2015

Teaching experience

Teaching assistant, Department of Biomedical Informatics (Columbia University)

Spring 2018

BINF G4002: Methods II: Computational Methods

Responsibility: Prepare and grade weekly programming labs and assignments, give lectures when the professor is absent, and hold weekly office hours.

Teaching assistant, Department of Biomedical Informatics (Columbia University) Fall 2018

BINF G4000: Acculturation to Programming and Statistics

Responsibility: Prepare and grade weekly programming labs and assignments, give lectures when the professor is absent, and hold weekly office hours.

Mentoring

Ali Turfah Master student Department of Statistics, Columbia University
Project: Developing novel deep learning-based natural language processing model for machine reading comprehension of literature. Summer 2020

Ananya Sistla, Jahnvi Shah Ardsley High School, New York
Project: Applying Natural Language Processing to structure medical evidence from PubMed. 2019-2020

Reviewing activities

Journals

Journal of the American Medical Informatics Association

Journal of Biomedical Informatics

Artificial Intelligence In Medicine

Conferences

IEEE International Conference on Healthcare Informatics

AMIA Annual Symposium

AMIA Informatics Summit

Editorial Committee

MEDINFO 2019