

Yining Hua

Boston, MA

yininghua@g.harvard.edu
+1 413-682-6681

Research Interest

AI for healthcare & social good, Psychiatric Epidemiology, Natural Language Processing

Education

Harvard T.H. Chan School of Public Health	Boston, MA
<i>Epidemiology (Ph.D.), advised by Prof. Lori Chibnik</i>	<i>Expected Jan 2028</i>
Harvard Medical School	Boston, MA
<i>Biomedical Informatics (MSc.), advised by Prof. Li Zhou</i>	<i>Sep. 2022 - Jan. 2023</i>
Harvard College	Cambridge, MA
<i>Undergraduate Visiting Program, advised by Dr. Chris Tanner</i>	<i>Sep. 2021 - May 2022</i>
Smith College	Northampton, MA
<i>Computer Science (B.A.), advised by Prof. Joseph O'Rourke</i>	<i>Sep. 2017 - May 2022</i>

Research Experience

Division of Internal Medicine, Brigham and Women's Hospital	Boston, MA
<i>Research Assistant</i>	<i>Oct. 2021 - Present</i>
<ul style="list-style-type: none"> Supervisors: Li Zhou, PhD; David W. Bates, MD; Lauren V. Moran, MD; Dinah Foer, MD Study COVID-19 infoveillance using large-scale longitudinal Twitter data. Develop multimodal algorithms to identify patient cohorts from EHRs. 	
Boston Children's Hospital	Boston, MA
<i>Research Assistant II</i>	<i>Jan. 2023 - August 2023</i>
<ul style="list-style-type: none"> Supervisors: Scott HS. Wang, MD, PhD; Michael Lingzhi Li, PhD Used language models to extract multi-class labels/outcomes from urodynamic investigation charts. Examined efficiency of different active learning algorithms in multi-task and multi-class prediction. 	
Department of Dermatology, Massachusetts General Hospital	Boston, MA
<i>Research Trainee</i>	<i>Nov. 2021 - May 2022</i>
<ul style="list-style-type: none"> Supervisors: Yevgeniy R. Semenov, MD, PhD Analyzed IBM Truven MarketScan data for health utilization studies. Examined image segmentation methods for melanoma stage II predictions. 	
Anesthesia, Beth Israel Deaconess Medical Center	Boston, MA
<i>Clinical Research Assistant II</i>	<i>Jun. - Sep. 2022</i>
<ul style="list-style-type: none"> Supervisors: Maximilian S. Schaefer, MD; Haobo Ma, MD Developed algorithms for identifying delirium patient cohorts. Increased accuracy of patient identification from ICD codes (25% to 86%). 	
CELEHS, Harvard Medical School	Boston, MA
<i>Research Intern</i>	<i>May - Dec. 2021</i>
<ul style="list-style-type: none"> Supervisors: Tianxi Cai, PhD Constructed cross-ontology hierarchical medical relations. Improved mapping accuracy by 30% over exact matching. 	

General Internal Medicine, Massachusetts General Hospital

Boston, MA

Research Intern

Jun. - Nov. 2021

- Supervisors: Jennifer S. Haas, MD
- Examined COVID-19's impact on the insurance status of breast cancer patients.

Computer Science, Smith College

Northampton, MA

Research Assistant

May 2019 - Aug. 2020

- Supervisors: Jamie C. Macbeth, PhD
- Developed an English Language Interpreter based on Conceptual Dependency theory.
- Compared performance of humans and pre-trained language models.

Publications & Preprints (denotes equal contribution)*

Liu, J.*, Zhou, P.*, **Hua, Y.***, et al. (2023). Benchmarking Large Language Models on CMExam—A Comprehensive Chinese Medical Exam Dataset. *Advances in Neural Information Processing Systems*, 36.

Zhou, P., Gao, J., Xie, Y., Ye, Q., **Hua, Y.**, Kim, S. (2022). Equivariant Contrastive Learning for Sequential Recommendation. *Proceedings of the ACM Web Conference (WWW)*, 2023.

Xie, Y., Gao, J., Zhou, P., Ye, Q., **Hua, Y.**, Kim, J., Wu, F., Kim, S. (2023). Rethinking Multi-Interest Learning for Candidate Matching in Recommender Systems. *Proceedings of the 17th ACM Conference on Recommender Systems - RecSys*, 2023.

Zeng, Q., Garay, L., Zhou, P., Chong, D., **Hua, Y.**, Wu, J., Pan, Y., Zhou, H., Voigt, R., Yang, J (2023). GreenPLM: Cross-lingual transfer of monolingual pre-trained language models at almost no cost. *the 32nd International Joint Conference on Artificial Intelligence*.

Hua, Y., Wang, L., Nguyen, V., Rieu-Werden, M., McDowell, A., Bates, D. W., Zhou, L (2023). A Deep Learning Approach for Transgender and Gender Diverse Patient Identification in Electronic Health Records. *Journal of Biomedical Informatics*.

Wu, J.*, Wang, L.*, **Hua, Y.**, Li, M., Zhou, L., Bates, D. W., Yang, J. (2023). Trend and Co-occurrence Network of COVID-19 Symptoms From Large-Scale Social Media Data: Infoveillance Study. *Journal of Medical Internet Research*, 25, e45419.

Wu, J., Wu, X., **Hua, Y.**, Lin, S., Zheng, Y., Yang, J. (2023). Exploring Social Media for Early Detection of Depression in COVID-19 Patients. *Proceedings of the ACM Web Conference (WWW)*, 2023.

Macbeth, J., Chang, E., Chen, G., **Hua, Y.**, Grandic, S., Zheng, W. X. (2023). A Broader Range for 'Meaning the Same Thing': Human Against Machine on Hard Paraphrase Detection Tasks. *Advances in Cognitive Systems*.

Ye, Q., Liu, J., Chong, D., Zhou, P., **Hua, Y.**, Liu, A. (2023). Qilin-med: Multi-stage knowledge injection advanced medical large language model. *arXiv preprint arXiv:2310.09089*.

Zhou, H., Gu, B., Zou, X., Li, Y., Chen, S. S., Zhou, P., Liu, J., **Hua, Y.**, Mao, C., Wu, X., Li, Z., Liu, F. (2023) A Survey of Large Language Models in Medicine: Progress, Application, and Challenge. *arXiv preprint arXiv:2311.05112*.

Liu, J.*, Wang, Z.*, Ye, Q.*, Chong, D.*, Zhou, P.*, **Hua, Y.*** (2023). Qilin-Med-VL: Towards Chinese Large Vision-Language Model for General Healthcare. *arXiv preprint arXiv:2310.17956*.

Guo, Z., **Hua, Y.** (2023). Continuous Training and Fine-tuning for Domain-Specific Language Models in Medical Question Answering. *arXiv preprint arXiv:2311.00204*.

Zhou, P., Cao, M., Huang, Y. L., Ye, Q., Zhang, P., Liu, J., Xie, Y., **Hua, Y.**, Kim, J. (2023). Exploring recommendation capabilities of gpt-4v (ision): A preliminary case study. *arXiv preprint arXiv:2311.04199*.

Zhou, P., Wang, Z., Chong, D., Guo, Z., **Hua, Y.**, Su, Z., Teng, Z., Wu, J., Yang, J. (2022). METS-CoV: A Dataset of Medical Entity and Targeted Sentiment on COVID-19 Related Tweets. *Advances in Neural Information Processing Systems*, 35, 21916-21932.

Li, M.*, **Hua, Y.***, Liao, Y., Zhou, L., Li, X., Wang, L., Yang, J. (2022). Tracking the Impact of COVID-19 and Lockdown Policies on Public Mental Health Using Social Media: Inveigilliance Study. *Journal of Medical Internet Research*, 24(10), e39676.

Jiang, H.*, **Hua, Y.***, Beeferman, D., Roy, D. (2022). Annotating the Tweepbank corpus on named entity recognition and building NLP models for social media analysis. *Proceedings of the 13th Language Resources and Evaluation Conference*, 28.

Hua, Y.*, Jiang, H.*, Lin, S., Yang, J., Plasek, J. M., Bates, D. W., Zhou, L. (2022). Using Twitter data to understand public perceptions of approved versus off-label use for COVID-19-related medications. *Journal of the American Medical Informatics Association*, 29(10), 1668-1678.

Wan, G., Nguyen, N., Liu, F., DeSimone, M. S., Leung, B. W., Rajeh, A., Collier, M. R., ..., **Hua, Y.**, et al. (2022). Prediction of early-stage melanoma recurrence using clinical and histopathologic features. *NPJ Precision Oncology*, 6(1), 79.

Conference Talks (excluding non-presenter sessions)

Conference on Neural Information Processing Systems (NeurIPS), New Orleans, LA. *Benchmarking Large Language Models on CMExam—A Comprehensive Chinese Medical Exam Dataset*. Dec 2023.

AMIA Annual Symposium, New Orleans, LA. *Streamlining Social Media Information Retrieval for Public Health Research with Deep Learning*. Nov 2023.

IEEE International Conference on Healthcare Informatics (ICHI), Houston, TX. *Streamlining Social Media Information Retrieval for Public Health Research with Deep Learning*. Jun 2023.

IEEE International Conference on Healthcare Informatics, Houston, TX. *A Deep Learning-driven Approach for Transgender and Gender Diverse Patient Identification in EHRs*. Jun 2023.

American Academy of Neurology Annual Meeting, Boston, MA. *Contemporary treatment and outcomes of myasthenia gravis in the United States*. Apr 2023.

American Medical Informatics Association (AMIA) Annual Symposium, Washington, D.C. *Identification of transgender and gender diverse individuals in electronic health records*. Nov 2022.

AMIA Annual Symposium, Washington, D.C. *Using Twitter to Understand Public Perceptions of COVID-19 Drugs*. Nov 2022.

Asia-Pacific Babylab Constellation Conference, Hong Kong. *Quantifying the bilingual (dis) advantage in vocabulary acquisition*. Dec 2021.

Teaching Experience

BMI 707/EPI-290: Deep Learning for Biomedical Data	Harvard Medical School
<i>Teaching Fellow</i>	<i>Spring 2023</i>
S-043/Stat-151: Multilevel and Longitudinal Models	Harvard School of Education
<i>Teaching Fellow</i>	<i>Summer 2021</i>
Machine Learning for Self-Driving Cars	Harvard School of Public Health
<i>Teaching Assistant</i>	<i>Summer 2020</i>
PHY 215: Light, Relativity, and Quantum Physics	Smith College
<i>Teaching Assistant</i>	<i>Spring 2020</i>
PHY 210: Math Methods of Physical Sciences and Engineering	Smith College
<i>Teaching Assistant</i>	<i>Fall 2019</i>

Awards & Accolades

The 11th IEEE ICHI	2023
<i>Best poster</i>	
The 36th Conference on NeurIPS	2022
<i>Spotlight long paper</i>	
The AMIA Annual Symposium	2022
<i>Distinguished poster</i>	
The Sigma Xi honor society	2021
<i>Nominated outstanding student researcher</i>	
Smith College	2017 - 2020
<i>Dean's List</i>	

Fundings & Fellowships

Harvard T.H Chan School of Public Health	2023
<i>The Brian and Heidi MacMahon Epidemiology Educational Fund</i>	
The 37th Conference on NeurIPS	2023
<i>Travel Fund</i>	
The Grace Hopper Celebration of Women in Computing	2020
<i>Student scholarship</i>	
Smith College	2020
<i>Project Coach fellowship</i>	
Smith College	2019
<i>Praxis fellowship</i>	

Review Experience

The Journal of American Medical Informatics Association (JAMIA)	2023
<i>Reviewed 3 articles</i>	
The AMIA Annual Symposium	2023
<i>Reviewed 2 articles and 3 abstracts</i>	
The AMIA Informatics Summit	2023
<i>Reviewed 2 articles and 3 abstracts</i>	

The International Journal of Medical Informatics (IJMI) <i>Reviewed 2 articles</i>	2022
The 29th International Conference on Computational Linguistics (COLING) <i>Reviewed 3 articles</i>	2022
The AMIA Annual Symposium <i>Reviewed 2 articles and 3 abstracts</i>	2022

Other Service and Activities

MIT Flying Club <i>Member</i>	Fall 2021 - Current
MIT Kickboxing Club <i>Member</i>	Fall 2021 - Current
Harvard Chinese Students and Scholars Association <i>Co-chair of the Seminar Department</i>	2021-2022
Harvard China Alumni Conference <i>Organizer and Moderator: Forum of Society and Culture</i>	2022
AI Time, Tsinghua University, China <i>Ph.D. Debate Series Moderator</i>	Fall 2020 - Summer 2021
Harvard Chinese Students and Scholars Association <i>Fellow of the Seminar Department</i>	2020-2021
The AEMES Scholars program, Smith College <i>STEM Mentor</i>	Fall 2020
The MIT Educational Studies Program (ESP) <i>Instructor</i>	Fall 2020
The Project Coach Program, Smith College <i>Academic Coach Fellow</i>	Spring 2020
APS Conferences for Undergraduate Women in Physics, UMass Amherst <i>Volunteer</i>	01/2019