Jie YANG, PhD, FACMI, FAMIA

CONTACT INFORMATION

1 Brigham Circle, Boston, MA, USA.
Work: +1 857-260-3191
Email: jyang66@bwh.harvard.edu

Google Scholar; *GitHub*
Website: https://ylab.top

RESEARCH INTERESTS

AI in healthcare; Clinical natural language processing; Large Language Models.

RESEARCH APPOINTMENTS

Harvard University, Boston, USA 2023-Present

Assistant Professor, Harvard Medical School

Lead Investigator, Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital

Affiliate faculty, Broad Institute of MIT and Harvard

Affiliate faculty, Harvard Data Science Initiative

Affiliate faculty, Kempner Institute of Harvard University

- Topic: *Artificial intelligence in healthcare, drug effectiveness and safety.*

Zhejiang University, Hangzhou, China 2020-2023

Assistant Professor (ZJU 100 Young Professor), School of Medicine

Adjunct Full Professor, the Second Affiliated Hospital of Zhejiang University

- Topic: *Artificial intelligence in healthcare and public health, precision medicine.*

Harvard Medical School, Boston, USA 2018-2020

Postdoctoral Research Fellow, Brigham and Women's Hospital

- Topic: *Artificial intelligence in medicine and healthcare, clinical NLP.*
- Host: *Prof. Li Zhou*

University of Oxford, Oxford, UK 2018-2018

Visiting Ph.D. student, Oxford-Man Institute of Quantitative Finance

- Topic: *Artificial Intelligence*
- Host: *Prof. Xiaowen Dong, Stefan Zohren*

EDUCATION

Singapore University of Technology and Design, Singapore 2014-2018

Ph.D., Computer Science

- Topic: *Deep learning for natural language processing*
- Advisor: *Prof. Yue Zhang*
- ***ISTD Best Dissertation Award***

University of Chinese Academy of Sciences, Beijing, China 2011-2014

M.S., Electronics

Chongqing University, Chongqing, China 2007-2011

B.S., IC Design and Integration System

B.S., Physics (double degree)

- Graduated from the special class for interdisciplinary study (top 0.5%)
- ***Outstanding Graduate Award***

Honors and Awards

- Fellow of American College of Medical Informatics (FACMI) 2025
- BWH DOM Chair's Research Award 2025
- Fellow of American Medical Informatics Association (FAMIA) 2025
- World Artificial Intelligence Conference (WAIC) Yunfan Award 2022
- Best Reviewer in CCL: *China National Conference on Computational Linguistics* 2020
- Kudos Reviewer in NAACL: *Association for Computational Linguistics* 2019
- Nvidia GPU Grant Award: *Nvidia Corp* 2019
- ISTD Best Dissertation Award: *Singapore University of Technology and Design* 2018
- COLING Best Paper Award: *Intl. Committee on Computational Linguistics* 2018
- ACL Best Demo Paper Award Nomination: *Association for Computational Linguistics* 2018
- Outstanding Reviewer in ACL: *Association for Computational Linguistics* 2018
- Outstanding Reviewer in COLING: *Intl. Committee on Computational Linguistics* 2018
- Best Technical Solution Award at A.I. Impact Weekend: *University of Oxford* 2018
- ACL Travel Award: *Association for Computational Linguistics* 2018
- President's Graduate Fellowship: *SUTD* 2014-2018
- Excellent Student Leader: *University of Chinese Academy of Sciences* 2014
- Outstanding Graduate : *Chongqing University* 2011
- Excellent Student (top 5%): *Chongqing University* 2010
- National Inspirational Scholarship: *Ministry of Education, PRC* 2008
- Honor Class (top 0.5%): *Chongqing University* 2007-2009

Grants

- **NIH-NLM R01 Grant** PI *$1.61M* 2025-2029
  Title: Developing large language models for drug safety and effectiveness causal analysis.
- **NIH-NIA R01 Grant** Multiple-PI (with Dr. Joshua Lin) *$3.69M* 2025-2030
  Title: Optimizing validity of comparative effectiveness research in Alzheimer's disease and related dementias using large language models.
- **FDA Sentinel Project** PI *$150K* 2024-2025
  Title: Incorporate a range of frequently used engineering features from EHRs into the Sentinel common data model in the Sentinel EHR and claims linked Data Partner network
- **Goldberg Scholarship** PI *$60K* 2024-2025
  Title: Efficient Phenotype Extraction from Electronic Health Records Using Large Language Models to Support Studies on Treatment Effects.
- **PCORI** Co-PI (with Dr. Richard Wyss) *$350K* 2024-2027
  Title: Transportability of Machine Learning-Enabled Confounding Control Methods Across EHR Databases. (LLM Methods Supplement Grant)
- **Brigham and Women's Hospital Fund To Sustain Grant** PI *$90K* 2024-2025
  Title: Developing large language models for drug safety and effectiveness causal analysis.
- **NVIDIA GPU Grant** PI V100 GPU 2019-2020
  Supported by the NVIDIA Corp.

Selected Publications

($*$ equal contribution, # corresponding/senior author. ACL/NeurIPS/IJCAI/COLING/NAACL/WWW are top-tier AI conferences, NEJM AI/JAMIA/JAMA Network Open are top-tier journals for healthcare or medical informatics.)

[1] Jiageng Wu, Bowen Gu, Ren Zhou, Kevin Xie, Doug Snyder, Yixing Jiang, Valentina Carducci, Richard Wyss, Rishi J Desai, Emily Alsentzer, Leo Anthony Celi, Adam Rodman, Sebastian Schneeweiss, Jonathan H. Chen, Santiago Romero-Brufau, Kueiyu Joshua Lin, **Jie Yang#**. BRIDGE: Benchmarking Large Language Models for Understanding Real-world Clinical Practice Text. *Nature Biomedical Engineering (under revision), 2025*

[2] Chong Shao, Douglas Snyder, Chiran Li, Bowen Gu, Kerry Ngan, Chun-Ting Yang, Jiageng Wu, Richard Wyss, Kueiyu Joshua Lin, **Jie Yang#**. Scalable Medication Extraction and Discontinuation Identification from Electronic Health Records Using Large

Language Models. *Journal of Clinical Epidemiology, 2025*

[3] Jiageng Wu, Xian Wu, Yefeng Zheng, **Jie Yang#**. Clinical Pathway-Aware Large Language Models for Reliable and Transparent Medical Dialogue. *Journal of Biomedical Informatics, 2025*

[4] Jiageng Wu, Richard Wyss, Kueiyu Joshua Lin, **Jie Yang#**. Cognitive Assessment of Dementia Patients Using Large Language Models. *Alzheimer's & Dementia (accepted), 2025*

[5] Jiageng Wu, Richard Wyss, Kueiyu Joshua Lin, **Jie Yang#**. Large Language Model for Predicting Altered Mental Status in Dementia Patients Using Electronic Health Records. *Alzheimer's & Dementia (accepted), 2025*

[6] Shixu Lin, Lucas Garay, Yining Hua, Zhijiang Guo, Wanxin Li, Minghui Li, Yujie Zhang, Xiaolin Xu, **Jie Yang#**. Analysis of Longitudinal Social Media for Monitoring Symptoms During a Pandemic. *Journal of Biomedical Informatics, 2025*

[7] Richard Wyss, **Jie Yang**, Sebastian Schneeweiss, Joseph M Plasek, Li Zhou, Thomas Deramus, Janick G Weberpals, Kerry Ngan, Theodore N Tsacogianis, Kueiyu Joshua Lin. Natural language processing for scalable feature engineering and ultra-high-dimensional confounding adjustment in healthcare database studies. *Journal of Biomedical Informatics, 2025*

[8] Wanxin Li, Yining Hua, Peilin Zhou, Li Zhou, Xin Xu, Jie Yang **Jie Yang#**. Analysis of Longitudinal Social Media for Monitoring Symptoms During a Pandemic. *Journal of Medical Internet Research, 2025*

[9] Jiageng Wu, Xiaocong Liu, Minghui Li, Wanxin Li, Zichang Su, Garay Lucas, Zhiyun Zhang, Yujie Zhang, Qingcheng Zeng, Jie Shen, Changzheng Yuan#, **Jie Yang#**. Clinical Text Datasets for Medical Artificial Intelligence and Large Language Models: A Systematic Review. *The New England Journal of Medicine AI (**NEJM AI**), 2024.* ***Included by 2024 AMIA Informatics Year in Review.***

[10] Bowen Gu, Rishi J Desai, Kueiyu Joshua Lin#, **Jie Yang#**. Probabilistic Medical Predictions of Large Language Models. ***npj Digital Medicine**, 2024*.

[11] Jiageng Wu, Xian Wu#, Zhaopeng Qiu, Minghui Li, Shixu Lin, Yingying Zhang, Yefeng Zheng, Changzheng Yuan#, **Jie Yang#**. Large Language Models Leverage External Knowledge to Extend Clinical Insight Beyond Language Boundaries. *Journal of the American Medical Informatics Association (**JAMIA**), 2024* ***Included by 2024 AMIA Informatics Year in Review.***

[12] Michael Matheny, **Jie Yang**, Joshua Smith, Colin Walsh, Mohammed Al-Garadi, Sharon Davis, Keith Marsolo, Daniel Fabbri, Ruth Reeves, Kevin Johnson, Gerald Dal Pan, Robert Ball, Rishi Desai. Enhancing post-marketing surveillance of medical products with large language models ***JAMA Network Open**, 2024*

[13] Yining Hua, Jiageng Wu, Shixu Lin, Minghui Li, Yujie Zhang, Dinah Foer, SIwen Wang, Peilin Zhou, **Jie Yang#**, Li Zhou#. Streamlining Social Media Information Retrieval for Public Health Research with Deep Learning. *Journal of the American Medical Informatics Association (**JAMIA**), 2024*

[14] Zhiyun Zhang, Yining Hua, Peilin Zhou, Shixu Lin, Minghui Li, Yujie Zhang, Li Zhou, Yanhui Liao, **Jie Yang#**. The sexual and gender-diverse face more health challenges during COVID-19: A large-scale social media analysis with natural language processing. *Health Data Science, 2024*

[15] Jiageng Wu∗, Xian Wu∗, **Jie Yang#**. Guiding Clinical Reasoning with Large Language Models via Knowledge Seeds. *International Joint Conferences on Artificial Intelligence (**IJCAI**), 2024*

[16] Xiaocong Liu, Jiageng Wu, An Shao, Wenyue Shen, Panpan Ye, Yao Wang, Juan Ye#, Kai Jin#, **Jie Yang#**. Uncovering Language Disparity of ChatGPT on Retinal Vascular Disease Classification: Cross-Sectional Study. *Journal of Medical Internet Research, 2024*

[17] Xiaocong Liu, Jiageng Wu, An Shao, Wenyue Shen, Panpan Ye, Yao Wang, Juan Ye, Kai Jin, **Jie Yang#**. Uncovering Language Disparity of ChatGPT in Healthcare: Non-English Clinical Environment for Retinal Vascular Disease Classification. *American Medical Informatics Association (**AMIA**) Annual Symposium, 2023*

[18] Qingcheng Zeng, Lucas Garey, Peilin Zhou, Dading Chong, Han Zhou, Jiageng Wu, Yikang Pan, Han Zhou, Rob Voigt, **Jie Yang#**. GreenLLM: Cross-Lingual Transfer of Monolingual Pre-Trained Language Models at Almost No Cost. *International Joint Conferences on Artificial Intelligence (**IJCAI**), 2023*

[19] Jinsong Tang, **Jie Yang**, Yi Liu, Xiaocong Liu, ..., Yanhui Liao. Efficacy of WeChat-based online smoking cessation intervention (‘WeChat WeQuit’) in China: a randomised controlled trial. ***EClinicalMedicine**, 2023*

[20] Jiageng Wu, Xian Wu#, Yining Hua, Shixu Lin, Yefeng Zheng, **Jie Yang#**. Exploring Social Media for Early Detection of Depression in COVID-19 Patients. *Proceedings of the ACM Web Conference (**WWW**), 2023*

[21] Zeqiang Wang, Yile Wang, Zhiyang Teng, **Jie Yang#**. YATO: Yet Another Deep Learning Based Text Analysis Open Toolkit. *Empirical Methods in Natural Language Processing (**EMNLP**), 2023*

[22] Jiageng Wu, Lumin Wang, Minghui Li, Yining Hua, Li Zhou, David Bates, **Jie Yang#**. Trend and co-occurrence network of COVID-19 symptoms through large-scale social media: Infoveillance Study. *Journal of Medical Internet Research, 2023*

[23] Peilin Zhou, Zeqiang Wang, Dading Chong, Zhijiang Guo, Yining Hua, Zichang Su, Zhiyang Teng, Jiageng Wu, **Jie Yang#**. METS-CoV: A Dataset of Medical Entity and Targeted Sentiment on COVID-19 Related Tweets. *Advances in Neural Information Processing Systems (**NeurIPS**), 2022*

[24] Minghui Li, Yining Hua, Yanhui Liao, Li Zhou, Xue Li, Ling Wang, **Jie Yang#**. Tracking the impact of COVID-19 and lockdown policies on public mental health using social media: an infoveillance study. *Journal of Medical Internet Research, 2022*

[25] Qingcheng Zeng, Dading Chong, Peiling Zhou, **Jie Yang#**. Low-resource Accent Classification in Geographically-proximate Settings: A Forensic and Sociophonetics Perspective. *INTERSPEECH, 2022*

[26] Yining Hua, Hang Jiang, Shixu Lin, **Jie Yang**, Joseph M. Plasek, David W. Bates, Li Zhou. Using Twitter Data to Understand Public Perceptions of Approved versus Off-label Use for COVID-19-related Medications . ***JAMIA**, 2022*

[27] **Jie Yang**, John Lian, Yen Po Chin, Liqin Wang, Anna Lian, George F. Murphy, Li Zhou. Assessing the Prognostic Significance of Tumor-Infiltrating Lymphocytes in Patients With Melanoma Using Pathologic Features Identified by Natural Language Processing. ***JAMA Network Open**, 2021*

[28] **Jie Yang**, Liqin Wang, Neelam A. Phadke, Paige G. Wickner, Christian M. Mancini, Kimberly G. Blumenthal, Li Zhou. Development and Validation of a Deep Learning Model for Detection of Allergic Reactions Using Safety Event Reports Across Hospitals. ***JAMA Network Open***, 2020.

[29] Xingchen Wan*, **Jie Yang***, Slavi Marinov, Jan-Peter Calliess, Stefan Zohren, Xiaowen Dong. Sentiment correlation in financial news networks and associated market movements. *Scientific Reports, 2021*

[30] Jennifer Hass, ..., **Jie Yang**, Li Zhou, Anna Tosteson. Multilevel follow-up of cancer screening (mFOCUS): Protocol for a multilevel intervention to improve the follow-up of abnormal cancer screening test results. *Contemporary Clinical Trials, 2021*

[31] **Jie Yang**, Liqin Wang, Neelam A. Phadke, Paige G. Wickner, Christian M. Mancini, Kimberly G. Blumenthal, Li Zhou. Development and Validation of a Deep Learning Model for Detection of Allergic Reactions Using Safety Event Reports Across Hospitals. ***JAMA Network Open***, 2020.

[32] **Jie Yang**, Mary B. Landrum, Li Zhou, Alisa Busch. Disparities in Outpatient Visits for Mental Health and/or Substance Use Disorders During the COVID-19 Surge and Partial Reopening in Massachusetts. *General Hospital Psychiatry*, 2020.

[33] **Jie Yang**, Liqin Wang, Neelam A. Phadke, Paige G. Wickner, Christian M. Mancini, Kimberly G. Blumenthal, Li Zhou. Deep Learning to Detect Allergy Events from Hospital Safety Reports. *American Medical Informatics Association 2020 Annual Symposium (**AMIA**)*, 2020

[34] Neelam A. Phadke, Li Zhou, Christian M. Mancini, **Jie Yang**, Liqin Wang, Paige Wickner, Xiaoqing Fu, Kimberly G. Blumenthal. Allergic Reactions in Two Academic Medical Centers. *Journal of General Internal Medicine*, 2020.

[35] Yue Zhang*, Yile Wang*, **Jie Yang***. Lattice LSTM for Chinese Sentence Representation. *IEEE Transactions on Audio, Speech and Language Processing (**TASLP**)*, 2020

[36] Courtney Diamond, Steven Atlas, Tin Dang, **Jie Yang**, Li Zhou, Sanja Percac-Lima, Amy Wint, Kimberly Harris, John Orav, Erica Breslau, Shoshana Hort, Anna Tosteson, Jennifer Haas, Adam Wright. Implementing an IT-Based Intervention to Improve Follow-up Rates of Abnormal Cancer Screening Results: the mFOCUS Trial. *American Medical Informatics Association Annual Symposium (**AMIA**)*, 2020

[37] **Jie Yang**, Yue Zhang, Shuailong Liang. Subword Encoding in Lattice LSTM for Chinese Word Segmentation. *Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL**)*, 2019

[38] Tom Korach, **Jie Yang**, Sarah Collins Rossetti, Kenrick Cato, Min-Jeoung Kang, Christopher Knaplund, Kumiko Schnock, Jose P. Garcia, Haomiao Jia, Jessica M. Schwartz, Li Zhou. Mining clinical phrases from nursing notes to discover risk factors of patient deterioration. *International Journal of Medical Informatics (**IJMI**)*, 2019.

[39] Hongming Wang, **Jie Yang**, Yue Zhang. From Gensis to Creole language: Transfer Learning for Singlish Universal Dependencies Parsing and POS Tagging. *ACM Transactions on Asian and Low-Resource Language Information Processing (**TALLIP**)*, 2019.

[40] **Jie Yang**, Shuailong Liang and Yue Zhang. Design Challenges and Misconceptions in Neural Sequence Labeling. *In Proceedings of the 27th International Conference on Computational Linguistics (**COLING**)*, Santa Fe, USA, 2018. **Best Paper Award.**

[41] Yue Zhang* and **Jie Yang***. Chinese NER Using Lattice LSTM. *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (**ACL**), Long*, Melbourne, Australia, 2018.

[42] **Jie Yang** and Yue Zhang. NCRF++: An Open-source Neural Sequence Labeling Toolkit. *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (**ACL**), Demonstration*, Melbourne, Australia, 2018.

[43] **Jie Yang**, Yue Zhang, Linwei Li, Xingxuan Li. YEDDA: A Lightweight Collaborative Text Span Annotation Tool. *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (**ACL**), Demonstration*, Melbourne, Australia, 2018. **Best Demonstration Paper Nomination.**

[44] **Jie Yang**∗, Yue Zhang∗ and Fei Dong. Neural Word Segmentation with Rich Pretraining. *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (**ACL**), Long*, Vancouver, Canada, 2017.

[45] Hongmin Wang, Yue Zhang, GuangYong Chan, **Jie Yang** and Hai Leong Chieu. Universal Dependencies Parsing for Colloquial Singaporean English. *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (**ACL**), Long*, Vancouver, Canada, 2017.

[46] **Jie Yang**, Yue Zhang and Fei Dong. Neural Reranking for Named Entity Recognition. *In Proceedings of the Recent Advances in Natural Language Processing (**RANLP**)*, Varna, Bulgaria, 2017.

[47] Fei Dong, Yue Zhang, **Jie Yang**. Attention-based Recurrent Convolutional Neural Network for Automatic Essay Scoring. *In Proceedings of the 21st Conference on Computational Natural Language Learning (**CoNLL**)*, Vancouver, Canada, 2017.

[48] **Jie Yang**, Zhiyang Teng, Meishan Zhang and Yue Zhang. Combining Discrete and Neural Features for Sequence Labeling. *In Proceedings of the 17th International Conference on Intelligent Text Processing and Computational Linguistics (**CICLing**)*, Konya, Turkey, 2016.

[49] Meishan Zhang, **Jie Yang**, Zhiyang Teng and Yue Zhang. LibN3L:A Lightweight Package for Neural NLP. *In Proceedings of the 10th International Conference on Language Resources and Evaluation (**LREC**)*, Portoro, Slovenia, 2016.

SELECTED OPEN-SOURCE TOOLS

★ *represents the number of GitHub stars.*

- **NCRF++** ★ 1900 *https://github.com/jiesutd/NCRFpp*
  A neural sequence labeling framework with flexible integration of word, character, and handcrafted features with CRF layer (e.g., Bi-LSTM-CRF with char CNN).
- **YEDDA** ★ 1100 *https://github.com/jiesutd/YEDDA*
  An efficient graphical toolkit for chunk/entity/event annotation. It supports shortcut annotation, command line annotation, system recommendation and result analysis.
- **LatticeLSTM** ★ 1800 *https://github.com/jiesutd/LatticeLSTM*
  A new Lattice LSTM structure that utilizes character and word information. It gives the best performance on many Chinese NER/word segmentation tasks.
- **RichWordSegmentor** ★ 150 *https://github.com/jiesutd/RichWordSegmentor*
  A state-of-the-art neural transition-based Chinese word segmentor with rich pretraining character representation (multi-task structure).
- **TAHV** ★ 350 *https://github.com/jiesutd/Text-Attention-Heatmap-Visualization*
  A simple visualization tool for attention on text in NLP tasks.

INVITED TALKS

Understanding LLM Capabilities on Real-World Clinical Data Through 39 Million Predictions
- *UMass Chan (1/2026)*

Benchmarking LLMs on Real-World Clinical Data
- *AIM-AHEAD AI Optimization Subcore) NLP WG (12/2025)*

BRIDGE: Benchmarking Large Language Models for Understanding Real-world Clinical Practice Text
- *Observational Health Data Sciences and Informatics(OHDSI) NLP WG (6/2025)*
- *Google Verily, Boston, USA (6/2025)*

Large Language Models in Electronic Health Record Understanding
- *International Society of Pharmacovigilance Seminar Boston, USA (12/2024)*
- *Brigham and Women's Hospital, Boston, USA (12/2024)*
- *Kempner Institute, Harvard University, Boston, USA (11/2024)*

Natural Language Processing in Clinical Text Analysis
- *ByteDance AI Lab, Shanghai, China (11/2021)*
- *Harvard University, Boston, USA (10/2021, remote)*
- *VMDT Conference, Guangzhou, China (10/2021)*
- *Westlake University,Hangzhou (4/2021)*

Clinical Text Analysis and Mining using Artificial Intelligence
- *Duke University, Durham, USA (2/2020)*
- *Boston College, Boston, USA (2/2020)*
- *Harvard Medical School / Brigham and Women's Hospital, Boston, USA (2/2020)*
- *Microsoft Research, Seattle, USA (4/2020, cancelled)*
- *Peking University, Beijing, China (4/2020)*
- *The University of British Columbia, Canada (6/2020)*

Natural Language Processing in Financial Analysis
- *Peking University, China (12/2019)*
- *Southwest University, Chongqing, China (12/2019)*

Deep Learning for Sequence Labeling
- *Partners Healthcare group, Boston, USA (11/2019)*

Design of NCRF++: The Open-source Text Labeling Toolkit
- *Westlake University, Hangzhou, China (11/2018)*

Recent Advances in Sequence Labeling
- *Heilongjiang University, Harbin, China (9/2018)*

Lattice LSTM for Chinese Sequence Labeling
- *Tencent AI Lab, Seattle, USA (2018)*

PROFESSIONAL SERVICE

Academic Leadership
- World Health Organization (WHO) Global Clinical Platform 2024-2026
  Member of Technical Advisory Group
- npj Digital Medicine 2025-
  Guest Editor for the collection of "Impact of Agentic AI on Care Delivery"
- npj Digital Medicine 2025-
  Associate Editor
- IEEE Transactions on Neural Networks and Learning Systems 2025-
  Associate Editor
- npj Health Systems 2025-
  Associate Editor

- American Medical Informatics Association Informatics Summit (AMIA) 2025
  Scientific Program Committee (SPC)
- Annual Meeting of the Association for Computational Linguistics (ACL) 2025
  Area Chair
- Empirical Methods in Natural Language Processing (EMNLP) 2021,2022,2024
  Area Chair
- International Conference on Computational Linguistics (COLING) 2022, 2024
  Area Chair
- China National Conference on Computational Linguistics (CCL) 2023
  Area Chair
- International Journal of Medical Informatics (IJMI) 2022-2024
  Editorial Board Member
- Health Care Science 2022-
  Editorial Board Member
- Empirical Methods in Natural Language Processing (EMNLP) 2021
  Session Chair
- American Medical Informatics Association Annual Symposium (AMIA) 2020
  Scientific Program Committee (SPC), similar to the area chair in AI conferences

Journal Reviewer

- *JAMA*
- *Nature Medicine*
- *Lancet Digital Health*
- *Proceedings of the National Academy of Sciences (PNAS)*
- *JAMA Network Open*
- *Nature Communications*
- *International Journal of Surgery*
- *Computational Linguistics (Standing reviewer)*
- *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*
- *Journal of the American Medical Informatics Association (JAMIA)*
- *Journal of Substance Abuse Treatment*
- *Journal of Medical Internet Research (JMIR)*
- *International Journal of Medical Informatics*
- *Journal of Artificial Intelligence Research (JAIR)*
- *IEEE Transactions on BIG DATA*
- *Artificial Intelligence Review*
- *Clinical and Translational Allergy*
- *Natural Language Engineering*
- *Neurocomputing*
- *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*
- *Computer Speech & Language*

Program Committee

- *Intl. CONF on Learning Representations* (ICLR)
- *American Medical Informatics Association Annual Symposium* (AMIA)
- *Intl. CONF on Machine Learning* (ICML)
- *Association for Computational Linguistics* (ACL) *Outstanding reviewer, 2018*
- *Association for Advancement of Artificial Intelligence* (AAAI)
- *ACM Intl. CONF on Information and Knowledge Management* (CIKM)
- *Intl. CONF on Computational Linguistics* (COLING) *Outstanding reviewer, 2018*
- *Empirical Methods in Natural Language Processing* (EMNLP)
- *Asia-Pacific chapter of ACL* (AACL)
- *Intl. Joint Conferences on Artificial Intelligence* (IJCAI)
- *North American Chapter of the ACL* (NAACL) *Kudos reviewer, 2019*
- *The SIGNLL CONF on Computational Natural Language Learning* (CONLL)

- *China National CONF on Computational Linguistics* (CCL) *Best reviewer, 2020*
- *Pacific Asia CONF on Language, Information and Computation* (PACLIC)
- *Intl. CONF on Natural Language Processing and Chinese Computing* (NLPCC)
- *Intl. CONF on Asian Language Processing* (IALP)
- *Intl. Joint CONF on Natural Language Processing* (IJCNLP)

STUDENTS

- **Harvard University**

Alice Yang (Graduate Student, Harvard)		2025-present
Enqi Liu (Graduate Student, Harvard)		2025-present
Kevin Xie (Undergraduate student, MIT)		2025-present
Joshua Zhang (Undergraduate student, Harvard)		2025-present
Jasmine Liu (Undergraduate student, Harvard)		2025
Doug Snyder (Graduate Student, Harvard)		2024-2025
Ryan Li (Graduate Student, Harvard)		2024-2025
Camille Shao (Graduate Student, Harvard)		2024-2025
Bowen Gu (Graduate Student, Harvard) (with Dr. Rishi Desai)		2023-2024
Vivian Shao (Graduate Student, Harvard) (with Dr. Rishi Desai)		2023-2024
Yining Hua (Graduate Student, Harvard)		2021-2023
Sara Zhang (Graduate Student, Harvard)		2020–2021
Sunny Mahesh (Graduate Student, Harvard)		2019
Harvey Chin (Graduate Student, Harvard)		2019

- **Zhejiang University**

Xiaocong Liu (PhD Student)	*Clinical natural language processing*	2022-2023
Jiageng Wu (PhD Student)	*Clinical natural language processing*	2021-2023
Minghui Li (Graduate Student)	*Healthcare big data*	2021-2023
Lucas Garey (Graduate Student)	*Clinical natural language processing*	2021-2023
Shixu Lin (Graduate Student)	*Big data in public health*	2021-2023
Wanxin Li (Graduate Student)	*AI in drug*	2022-2023
Zhiyun Zhang (Chu Kochen Honors College)	*AI in healthcare*	2021-2023
Zichang Su (Chu Kochen Honors College)	*AI in healthcare*	2021-2023

- **Weston High School, MA** (High school)

John Lian (High school)	*Pathology reports analysis for melanoma patients*	2019–2021

- **Singapore University of Technology and Design**

Linwei Li (Undergraduate)	*Deep learning in NLP*	2018
Xingxuan Li (Undergraduate)	*Deep learning in NLP*	2018

TEACHING EXPERIENCE

Harvard University
- Foundations of Clinical Data and its Applications. (*Guest speaker*) 2025

Zhejiang University
- Big Data in Health Science. (*Instructor*) 2021-2022
- Global Digital Health. (*Instructor*) 2023

Singapore University of Technology and Design, Singapore
- Computer System Engineering. (*TA with Prof. David Yau*) 2015-2016
- Machine Learning. (*TA with Prof. Wei Lu, Prof. Alexander Binder*) 2015-2016