Jie YANG, Ph.D.

CONTACT INFORMATION

Zijin'gang Campus, Zhejiang University, Hangzhou, China.
Work: +86 571 87077982
Email: y@zju.edu.cn
Google Scholar; *GitHub*
Website: https://ylab.top

RESEARCH INTERESTS

Artificial intelligence in healthcare; Clinical big data analysis; Applications of natural language processing & machine Learning.

RESEARCH APPOINTMENTS

Zhejiang University, Hangzhou, China 2020-Present

Assistant Professor (ZJU100 Young Professor), School of Medicine/Public Health

Adjunct 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., Microelectronics

- Advisor: *Professor Chao Zhao*
- ***Outstanding Dissertation Award***

Chongqing University, Chongqing, China 2007-2011

B.S., Electronics (honor class)

B.S., Physics (double degree)

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

HONORS AND AWARDS

- 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
- NO. 1 Individual Python Developer with Stars in Singapore: *git-awards.com* 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

PUBLICATIONS

Working Papers: *(∗ equal contribution, # corresponding author)*

- Minghui Li, Yining Hua, Yanhui Liao, Li Zhou, Xue Li, Ling Wang, **Jie Yang#**. Tracking the Impact of COVID-19 and the Lockdown Policy on Public Mental Health Using Social Media. *Under review, 2022*
- Haisheng Li∗, Jiageng Wu∗, Yumei Zhang, Wei Luo, Ning Li, Zhengzheng Zhang, Shixu Lin, Xinggang Wang, Chunmao Han, Zhiqiang Yuan, Gaoxing Luo#, **Jie Yang#**. Development and Validation of a Machine Learning based Decision Support System for Tracheostomy in Burn Patients. *Under Review, 2022*
- Jiageng Wu∗, Lumina Wang∗, Minghui Li, Yining Hua, Li Zhou, **Jie Yang#**. Mining Clinical Characteristics of COVID-19-Related Symptoms Using Social Media. *Under Review, 2022*
- Qingcheng Zeng∗, Lucas Garey∗, Peilin Zhou∗, Dading Chong, Han Zhou, Jiageng Wu, Yikang Pan, **Jie Yang#**. Translating Large Pre-trained Language Models to Low Resource Languages with (almost) No Cost. *Under Review, 2022*
- Zeqiang Wang∗, Yile Wang∗, Zhiyang Teng, **Jie Yang#**. YATO: Yet Another Deep Learning Based Text Analysis Open Toolkit. *Under Review, 2022*

Published Papers: *(ACL/COLING/NAACL are top-tier artificial intelligence conferences)*

[1] 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*

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

[3] 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*

[4] **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*

[5] Liqin Wang, John Laurentiev, **Jie Yang**, Yen Po Chin, Rebecca E. Amariglio,; Deborah Blacker,Reisa A. Sperling, Gad A. Marshall, Li Zhou. Development and Validation of a Deep Learning Model for Earlier Detection of Cognitive Decline From Clinical Notes in Electronic Health Records. *JAMA Network Open, 2021*

[6] 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*

[7] 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*

[8] **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.

[9] **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.

[10] **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

[11] 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.

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

[13] 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 2020 Annual Symposium (**AMIA**)*, 2020

[14] **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

[15] 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.

[16] 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.

[17] **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.**

[18] 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.

[19] **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.

[20] **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.**

[21] **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.

[22] 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.

[23] **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.

[24] 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.

[25] **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.

[26] 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++** ★ 1800 *https://github.com/jiesutd/NCRFpp*
  A neural sequence labeling framework with flexiable integration of word, character, and handcrafted features with CRF layer (e.g. Bi-LSTM-CRF with char CNN).
- **YEDDA** ★ 800 *https://github.com/jiesutd/YEDDA*
  An efficient graphical toolkit for chunk/entity/event annotation. It supports shortcuts annotation, command line annotation, system recommendation and result analysis.
- **LatticeLSTM** ★ 1500 *https://github.com/jiesutd/LatticeLSTM*
  A new Lattice LSTM structure which 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 neural rich pretraining character representation (multi-task structure).
- **TAHV** ★ 250 *https://github.com/jiesutd/Text-Attention-Heatmap-Visualization*
  A simple visualization tool for attentions on text in NLP tasks.

GRANTS

- **Data Driven Diagnosis-related Group** PI *30,000 USD* 2021-2022
  Supported by the Research Institute of Health Insurance of Zhejiang Province.
- **Knowledge Graph based QA System for Healthy Cooking** PI *30,000 USD* 2022-2023
  Supported by the MIDIA GROUP.
- **AI based Psychiatric Diagnosis and Phenotype Extraction** PI *45,000 USD* 2022-2023
  Supported by the ALIBABA GROUP.

- **NVIDIA GPU Grant** PI 2019-2020
  Supported by the NVIDIA Corp.

INVITED TALKS

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
- International Journal of Medical Informatics (IJMI) 2022-
  Editorial Board Member
- Health Care Science 2022-
  Editorial Board Member
- Key Laboratory of Intelligent Preventive Medicine of Zhejiang Province 2022-
  Director of Artificial Intelligence
- Empirical Methods in Natural Language Processing (EMNLP) 2021,2022
  Area Chair
- International Conference on Computational Linguistics (COLING) 2022
  Area Chair
- 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 Network Open*
- *Nature Communications*
- *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

- **Zhejiang University**

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

- **Harvard University**

Yining Hua (Graduate Student)	*Clinical natural language processing*	2021-
Sara Zhang (Graduate Student)	*Clinical natural language processing*	2020–2021
Sunny Mahesh (Graduate Student)	*Clinical natural language processing*	2019
Harvey Chin (Graduate Student)	*Clinical natural language processing*	2019

- **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

Zhejiang University, China

- Big Data in Health Science. (*Instructor*) 2021

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

REFERENCES

Li Zhou, M.D., Ph.D., **Harvard Medical School.** (lzhou@bwh.harvard.edu)
- Associate Professor, Brigham and Women's Hospital
- ⋆ *Dr. Zhou is my close collaborator and was my postdoctoral host .*

Christine G. Lian M.D., **Harvard Medical School.** (cglian@bwh.harvard.edu)
- Associate Professor, Dana-Farber Cancer Institute
- ⋆ *Dr. Lian is my close collaborator on a skin-cancer project.*

Yue Zhang Ph.D., **Westlake University.** (yue.zhang@wias.org.cn)
- Professor, School of Engineering
- ⋆ *Dr. Zhang was my PhD supervisor when he worked at SUTD, Singapore.*

Kimberly Blumenthal, M.D., **Harvard Medical School.** (kblumenthal@mgh.harvard.edu)
- Assistant Professor, Massachusetts General Hospital
- ⋆ *Kim is my close collaborator on deep learning based allergy research.*