

Junyu Luo | Curriculum Vitae

(Pennsylvania, State College) USA

✉ luoshuiti@outlook.com • 🌐 soap117.github.io/junyu.github.io/

Education & Awards

Academic Qualifications

- **Pennsylvania State University** **Ph.D.**
Information Sciences and Technology, GPA: 3.91/4.00 2020.08–2024
- **Sichuan University** **Bachelor**
Computer Science, GPA: 3.86/4.00, Major GPA: 3.924/4.00 2020.06

Honors & Awards

- **The Award of Excellence, MSRA Internship Program** 2020
- **National Scholarship** (THREE TIMES) 2015-2018
- **The First Prize Scholarship of Sichuan University** (THREE TIMES) 2015-2018
- **The first prize in Sichuan Province Lanqiao Programming Contest** 2017
- **Model Student of Academic Records, Sichuan University** 2017
- **The first prize in Sichuan University Mathematics Competition** 2016
- **The second prize in Sichuan University ACM Programming Contest** 2016

Research and Work Experience

- **Research Assistant on Machine Learning for Healthcare** **Dr. Fenglong Ma**
Pennstate University IST, Pennsylvanian, USA Feb 2020–Now
 - **Multi-modality Pre-training of EHR Data**
Paper: Hierarchical Pretraining on Multimodal Electronic Health Records.
TLDR: A new multi-modality a novel, general, and unified pretraining framework called MEDHMP for multi-modality health data.
 - **Automatic ICD Coding based on Diagnosis Text**
Paper: Fusion: Towards Automated ICD Coding via Feature Compression.
Using information compression to reduce the noise and improve speed.
<https://aclanthology.org/2021.findings-acl.184.pdf>
Paper: CoRelation: Boosting Automatic ICD Coding Through Contextualized Code Relation Learning.
TLDR: Improve performance through modeling contextualized code relations through graph network.
 - **Medical Text Simplification**
Paper: Benchmarking Automated Clinical Language Simplification: Dataset, Algorithm, and Evaluation.
TLDR: Designing a controllable simplification pipeline for using external medical dictionary knowledge.
<https://aclanthology.org/2022.coling-1.313.pdf>
 - **Electric Health Record Mining**
Paper: HiTANet: Hierarchical Time-Aware Attention Networks for Risk Prediction on Electronic Health Records.
TLDR: Using two-level transformers to model the complex EHR code sequential data.
https://dl.acm.org/doi/pdf/10.1145/3394486.3403107?casa_token=HLY2Uo1-v5kAAAAA:v70yG0bR1jJM1oMJMEpXvKk0LiV7jLgRpNFZjrDfwQvsNB52m_AafEfmbFxD5iPM3pgR0vwBpNQ
- **Research Intern on Natural Language Processing** **Dr. Danica Xiao**
Relativity, USA June 2023–Aug 2023
 - **Designing Algorithm for Preventing Hallucination for Large Language Models (LLMs).**
Paper: Zero-Resource Hallucination Prevention for Large Language Models (On submission)
TLDR: Using prompt engineering to perform self-evaluation under the zero-resource setting to test the understanding of LLMs to the instructions.

- Research Intern on Machine Learning for Clinical Data**
IQVIA, USA
 - **Designing Clinical Trial Retrieval Algorithm Based on Trial Protocols.**
Paper: Clinical Trial Retrieval via Multi-grained Group-based Similarity Learning
TLDR: *Designing hierarchical matching model for trial protocols with novel group-based training loss and 2D word matching.*
 - **Designing Personalized Drug Risk Prediction Model.**
Paper: pADR: Towards Personalized Adverse Drug Reaction Prediction by Modeling Multi-sourced Data.
TLDR: *Incorporating the patient's EHR modality with the drug molecular level information.*
- Research Intern on Knowledge Computing**
Microsoft Research Lab - Asia (MSRA), Beijing, China
 - **Automatic Pattern Recognition from Power Point Design.**
TLDR: *Transforming the pattern matching into a sequential matching problem.*
 - **Object Detection for Special Chart Images.**
Paper: ChartOCR: Data Extraction from Charts Images via a Deep Hybrid Framework
TLDR: *High precision point based object detection for chart objects.*
https://dl.acm.org/doi/pdf/10.1145/3357384.3358104?casa_token=20Jse6NYBnAAAAAA:ftUVD1L6-uQMklibfnkaKqre0Ueo7VXv-BTVs-vDEX4MU1MvS1IWTLYOANsgU9X0CbVLn-YRsr4
Paper: Hybrid Cascade Point Search Network for High Precision Bar Chart Component Detection
TLDR: *High precision object detection through cascade updating.*
<https://ieeexplore.ieee.org/abstract/document/9412144>
- Research Intern on Medical Images**
University of California(UCLA), Los Angeles, USA
 - Selected as a CSST Intern under guidance of Professor William Hsu of Medical Imaging Informatics Lab
 - Built a pipeline system for pulmonary nodule analysis from the raw CT images using deep learning algorithms
 - Assisted with data preprocessing and algorithms optimization
- Research Intern on Natural Language Processing**
Shenzhen Institutes of Advanced Technology(SIAT), Shenzhen, China
 - **Developed methods to generate semantic embedding for long sentences and cross-model searching**
Paper: Cross-modal Image-Text Retrieval with Multitask Learning.
TLDR: *Using back-encoding to ensure the cross-modality relation.*
<https://dl.acm.org/doi/10.1145/3357384.3358104>
- Research Intern on Deep Learning**
MI LAB Sichuan University(SCU), Chengdu, China
 - Finished one National Training Program of Innovation as leader and major developer and one independent research program under guidance of Professor Jianchen Lv

Dr. Cheng Qian

May 2022–Dec 2022

Dr. Jinpeng Wang

Mar 2019–Jan 2020

Dr. William Hsu

Jul 2018–Sep 2018

Dr. Min Yang

Sep 2017–Jul 2018

Dr. Jianchen Lv

Sep 2016–Jul 2017

Skills

- Experience in dealing with different kinds of data (Image, Text, Video, Web, Audio, CT)
- Experience in all kinds of deep learning frameworks, including transformers, LLMs, diffusion, GAN, graph neural networks, information retrieval frameworks, and object detection frameworks
- Experience in building web pages and mobile applications
- Master in Python (PyTorch, TensorFlow, Keras) and familiar with C, C++, Java, JavaScript

Publications

Tutorials.....

- Fenglong Ma, Muchao Ye, **Junyu Luo**, Cao Xiao, and Jimeng Sun. *Advances in Mining Heterogeneous Healthcare Data*. Conference Tutorial at the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2021.

Conferences & Journals.....

- **Junyu Luo**, Cheng Qian, Xiaochen Wang, Lucas Glass, and Fenglong Ma. 2023. *pADR: Towards Personalized*

- Adverse Drug Reaction Prediction by Modeling Multi-sourced Data*. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM 23), October 21–25, 2023, Birmingham, United Kingdom.
- **Junyu Luo**, Zhi Qiao, Lucas Glass, Cao Xiao, and Fenglong Ma. 2023. *ClinicalRisk: A New Therapy-related Clinical Trial Dataset for Predicting Trial Status and Failure Reasons*. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM 23), October 21–25, 2023, Birmingham, United Kingdom.
 - **Junyu Luo**, Junxian Lin, Chi Lin, Cao Xiao, Xinning Gui and Fenglong Ma. *Benchmarking Automated Clinical Language Simplification: Dataset, Algorithm, and Evaluation*. Proceedings of the 29th International Conference on Computational Linguistics (COLING 2022), OCTOBER 12-17, 2022, GYEONGJU, REPUBLIC OF KOREA.
 - **Junyu Luo**, Cao Xiao, Lucas Glass, Jimeng Sun and Fenglong Ma. *Fusion: Towards Automated ICD Coding via Feature Compression*. Findings of the 59th Annual Meeting of the Association for Computational Linguistics (Findings of ACL), 2021.
 - **Junyu Luo**, Zekun Li, Jinpeng Wang, Chin-Yew Lin: *ChartOCR: Data Extraction from Charts Images via a Deep Hybrid Framework*. Proceedings of the 2021 Winter Conference on Applications of Computer Vision (WACV), 2021.
 - **Junyu Luo**, Muchao Ye, Cao Xiao, Fenglong Ma. *HiTANet: Hierarchical Time-Aware Attention Networks for Risk Prediction on Electronic Health Records*. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2020.
 - **Junyu Luo**, Jinpeng Wang, and Chin-Yew Lin. *Hybrid Cascade Point Search Network for High Precision Bar Chart Component Detection*. Proceedings of the 25th International Conference on Pattern Recognition (ICPR), 2020.
 - **Junyu Luo**, Ying Shen, Xiang Ao, Zhou Zhao, Min Yang. *Cross-modal Image-Text Retrieval with Multitask Learning*. Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM), 2019.
 - **Junyu Luo**, Min Yang, Ying Shen, Qiang Qu, Haixia Chai. *Learning Document Embeddings with Crossword Prediction*. Proceedings of the Thirty-third AAAI Conference on Artificial Intelligence (AAAI), 2019.
 - Zhile Jiang, Shuai Yu, Qiang Qu, Min Yang, **Junyu Luo**, Juncheng Liu. *Multi-task Learning for Author Profiling with Hierarchical Features*. WWW (Companion Volume) 2018: 55-56.
 - **Junyu Luo**, Yong Xu, Chenwei Tang, Jiancheng Lv. *Learning Inverse Mapping by AutoEncoder Based Generative Adversarial Nets*. ICONIP (2) 2017: 207-216.
 - Suhan Cui, **Junyu Luo**, Muchao Ye, Jiaqi Wang, Ting Wang and Fenglong Ma. *MedSkim: Denoised Health Risk Prediction via Skimming Medical Claims Data*. Proceedings of the 22nd IEEE International Conference on Data Mining (ICDM 2022), Nov 28 - Dec 1, 2022, Orlando, FL.
 - Muchao Ye, Suhan Cui, Yaqing Wang, **Junyu Luo**, Cao Xiao, Fenglong Ma. *MedRetriever: Target-Driven Health Risk Prediction via Retrieving Unstructured Medical Text*. Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM), 2021.
 - Muchao Ye, Suhan Cui, Yaqing Wang, **Junyu Luo**, Cao Xiao, Fenglong Ma. *MedPath: Augmenting Health Risk Prediction via Medical Knowledge Paths*. Proceedings of the 30th The Web Conference (WWW), 2021.
 - Muchao Ye, **Junyu Luo**, Cao Xiao, Fenglong Ma. *LSAN: Modeling Long-term Dependencies and Short-term Correlations with Hierarchical Attention for Risk Prediction*. Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM), 2020.
 - Changqin Huang, Jia Zhu, Yuzhi Liang, Min Yang, Gabriel Pui Cheong Fung, **Junyu Luo**. *An efficient automatic multiple objectives optimization feature selection strategy for internet text classification*. Int. J. Mach. Learn. Cybern. 10(5): 1151-1163, 2019.

Submissions.....

- *Zero-Resource Hallucination Prevention for Large Language Models*. AAAI 2024.
- *CoRelation: Boosting Automatic ICD Coding Through Contextualized Code Relation Learning*. EMNLP 2023.
- *Clinical Trial Retrieval via Multi-grained Group-based Similarity Learning*. SDM 2024.