Junyu Luo

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**
- **The First Prize in Sichuan University Mathematics Competition** — **2016**

Skills

- Experience in processing different kinds of data (Image, Text, Web Data, Audio).
- Experience in all kinds of deep learning frameworks, including Transformers, LLMs, diffusion models, GAN, graph neural networks, information retrieval frameworks, and object detection frameworks.
- Experience in Natural Language Processing and Computer Vision related topics.
- Experience in building web pages and mobile applications for machine learning models.
- Master in Python (PyTorch, TensorFlow, Keras) and familiar with C#, C++, Java, and JavaScript.

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**
    Summary*: Developing a novel, multi-modal, and unified pretraining framework called MEDHMP for multi-modality health data pre-training.*
    Used Skills: Python, Multi-modality, Machine Learning, Spark
  - **Automatic ICD Coding based on Diagnosis Text**
    Summary*: Using information compression to reduce the clinical note noise and improve the speed of automatic ICD coding.*
    Used Skills: Transformers, NLP, ICD Coding
    Summary*: Improving ICD coding performance through modeling contextualized code relations through graph network.*
    Used Skills: Python, NLP, Machine Learning, Graph Network
  - **Medical Text Simplification**
    Summary*: Designing a controllable medical term simplification pipeline for using external medical dictionary knowledge.*
    Used Skills: Python, Neural Network Pipeline, NLP, Question Answering
  - **Electric Health Record Mining**
    Summary*: Using two-level transformers to model the complex EHR code sequential data to predict future diseases.*
    Used Skills: Python, Transformers, Disease Prediction
- **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).**
    Summary*: Using prompt engineering to perform self-evaluation under the zero-resource setting to test the understanding of LLMs to the instructions.*
    Used Skills: Python, NLP, Large Language Models, Prompt Engineering

- **Research Intern on Machine Learning for Clinical Data** — **Dr. Cheng Qian**
  IQVIA, USA — *May 2022–Dec 2022*
  - **Designing Clinical Trial Retrieval Algorithm Based on Trial Protocols.**
    Summary*: Designing hierarchical matching model for trial protocols with novel group-based training loss and 2D word matching.*
    Used Skills: Python, Machine Learning, NLP, Transformers, Information Retrieval
  - **Designing Personalized Drug Risk Prediction Model.**
    Summary*: Incorporating the patient's EHR modality with the drug molecular level information to predict the potential adverse reaction.*
    Used Skills: Python, Spark, Machine Learning
- **Research Intern on Knowledge Computing** — **Dr. Jinpeng Wang**
  Microsoft Research Lab - Asia (MSRA), Beijing, China — *Mar 2019–Jan 2020*
  - **Automatic Pattern Recognition from Power Point Design.**
    Summary*: Transforming the pattern matching into a sequential matching problem to discover potential design patterns.*
    Used Skills: C#, Sequential Matching
  - **Object Detection for Special Chart Images.**
    Summary*: Designing a high precision point-based object detection model for chart objects with web-based API.*
    Used Skills: Python, C#, Django, Web , Computer Vision, Object Detection
- **Research Intern on Natural Language Processing** — **Dr. Min Yang**
  Shenzhen Institutes of Advanced Technology(SIAT), Shenzhen, China — *Sep 2017–Jul 2018*
  - **Developed methods to generate semantic embedding for long sentences and cross-model searching**
    Summary*: Using back-encoding to ensure the cross-modality relation between learned text and image embeddings.*
    Used Skills: Python, Machine Learning, Information Retrieval