Hongye Liu

hongye.liu20@imperial.ac.uk | +86 131-4310-5769 | China

EDUCATION

Beijing Institute of Technology **Visiting Student** **01/2023 – Present**
- Direction: Artificial Intelligence in Speech (Healthcare area)
- Collaboration supervisor: Kun Qian / Bjoern Schuller

Imperial College London **MSc in Artificial Intelligence** **10/2021 – 10/2022**
- Average score: 69.87/100 (High Merit)
- Courses include: Deep Learning: 78.36/100;Reinforcement Learning: 66.17/100;Computational Optimisation: 85.1/100.
- Group project: Drug discovery with **Graph Neural Networks**. (Supervised by: Paul Bilokon)
- Individual project: Emotion Recognition in **Speech and further Modalities** (Supervised by: Bjoern Schuller)

Beijing Institute of Technology, Zhuhai **BSc in Applied Statistics (Data Science)** **09/2016 – 06/2020**
- GPA: 3.74 (Ranking: 1%) (Equal to First Class Honours)
- Courses include: Python Programing: 91/100; Bayesian Statistics: 97/100; Qualitative Data Analysis: 98/100; Non-Parametric Statistics: 97/100; Data Mining: 95/100; Experimental Design: 97/100; Machine Learning: 96/100.

PUBLICATIONS

[1] Mingzhou Xu, Longyue Wang, Derek F. Wong, **Hongye Liu**, Linfeng Song, Lidia S. Chao, Shuming Shi, and Zhaopeng Tu. A Benchmark for Zero Pronoun Recovery and Translation. **EMNLP 2022**.

[2] Y. Shen, **Hongye Liu** et al., Intelligent Flaw Detection of X-ray Images Based on Deep Learning. (Springer Science and Business Media B.V., 2021), vol. 105, pp. 558–566.

[3] A Method and Device of Bilingual Data for Web Fiction Translation via Self-Traning (Mainland & Hong Kong patent)

[4] A Method and Device for Automatic Alignment and Automatic Translation of Web Fiction (Mainland patent)

[5] A Method and Device of Large-scale Data Filtering for Machine Translation (Mainland patent)

[6] A Ray Defect Detection Method based on Mask R-CNN model (Mainland patent)

INTERNSHIPS

Tencent AI Lab **Engineering & Research Intern at NLP Group** **Shenzhen** **09/2020 – 09/2021**

Webnovel Translation
- Extracted CH-EN Webnovel data and optimized corpus alignment algorithm for the omission.
- Built domain corpus selections, back-translation and NMT model training pipeline.
- Developed front-end web and launch the translation engine. (Link: longyuewang.com/nmt)

Large-scale Parallel Corpus Filtering
- Integrated N-gram and LM to filter high-quality corpus and used Moore-lewis and FDA to select out related field data.

Zero pronoun Resolution
- Built test sets for zero pronoun translation from alignment, ZP labels and Part-of-Speech in four different domains.
- Assisted statistical data distribution, experiment and error analysis.

Beijing Genomics Institute (BGI) **AI Algorithm Intern at Big Data Center** **Shenzhen** **09/2019 – 01/2020**

NLP-AI Medical
- Designed a data pipeline based on electronic medical records to connect the upstream and downstream of the hospital.
- Built word segmentation, NER, relationship extraction to knowledge graph and knowledge query algorithm pipeline.

Biological gene-AI platform
- Investigated related papers and reproduced their work.
- Utilized Docker to package the models as mirror images and deployed them on the platform.

Integrated Design Project
- Build AutoEncoder, Resnet, and DeepFm models with Pytorch for feature extraction.
- Carry out cross-omics academic research based on real biological genetic data.

Search Engine for Scientific Research
- Crawled biomedical articles including gene data on the Internet.
- Deployed Lucene engine and built indexes to queried related articles.

PROJECTS

Emotion Recognition in Speech and further Modalities   **Individual Research Project**   **02/2022 – 10/2022**
- Pre-trained an audio model and visual model with extensively-used features.
- Proposed an audio-visual model comprised of Wav2vec in audio and Vision Transformer in visual parts.
- Reached the SOTA performance of the 3rd ABAW challenge

Drug Discovery With Graph Neural Networks   **Algorithm & Software**   **02/2022 – 04/2022**
- Predict the water solubility and toxicity of a molecule with the Graph Neural Network.
- Developed a front-end web with Streamlit. (Link: https://icl-drug-discovery.herokuapp.com/)
- Built an automatic DevOps pipeline with Docker and GitLab CI/CD.

A Social Platform for Chinese Oversea Student   **Project Leader**   **01/2021 – Present**
- Established a team consisting of technology, design, and marketing group.
- Investigated product-market fit (PMF) and carried out minimum viable product (MVP).
- Built a brand around the user group.
- Coordinated division of labour and cooperation between the different groups to build the product.

Application of Deep Learning on X-ray testing   **Project Leader**   **05/2019 – 05/2020**
- Collected the industrial data, designed AI algorithm for X-ray testing detection and achieved 81.3% recognition rate.
- Presented an academic report in the 2nd **Guangdong-Hong Kong- Macao Greater Bay Area** NDT forum.

Application of GAN in Seismic Image Recognition of Salt Mine   **Project Leader**   **03/2019 – 05/2019**
- Implemented image segmentation using the **GAN** model with Pytorch to recognize salt mine seismic images.
- Won the national second prize and presented an academic report at Sun Yat-sen University.

AWARDS

- Meritorious Winner at Mathematical Contest in Modeling (MCM) in 2018.
- 2nd Prize at The First China Geoscience Big Data Mining and AI Challenge in 2019.
- 3rd Prize at The MathorCup University Mathematical Modeling Challenge in 2018.
- Awarded Outstanding Student and Outstanding League Member 3 times since 2016

ACTIVITIES

- Vice President at the School Mathematical Modeling Association   **09/2017 – 07/2018**
- Assistant at the University Student Affairs Center   **09/2016 – 11/2017**
- Volunteer at the Guangdong QuanAi Social Service Center   **07/2018 – 09/2018**