

MINGXUAN LU

github.com/LouieBHLu | (914) 608-9291 | ml4799@columbia.edu | www.linkedin.com/in/lumingxuan/

New York, NY 10025 | Open to Remote and Relocating

EDUCATION

COLUMBIA UNIVERSITY

Master of Science in Electrical Engineering

Relevant Coursework: Big Data Analytics, Introduction to Databases, Reinforcement Learning

New York, NY

Expected Dec 2023

UM-SJTU JOINT INSTITUTE, SHANGHAI JIAO TONG UNIVERSITY

Bachelor of Engineering in Electrical and Computer Engineering

Relevant Coursework: Operating Systems, Methods and Tools for Big Data, Programming & Elem. Data Structures

Shanghai, China

Jun 2022

WORK EXPERIENCE

DOCTOR HEALTHX

Computer Vision Engineer

Shanghai, China

Apr 2022 - Present

- Served as one of the **founding engineers** to accelerate digital transformation for healthcare industry; received **HealthX Overachievement Award** for outstanding contributions
- Designed a **two-layer attention** convolutional neural network (CNN) to classify patients with placenta accreta spectrum (PAS) based on MRI images with an average accuracy (82%), **exceeding** that of physicians' diagnosis (67%)
- Equipped an Optical Coherence Tomography (OCT) instruments company with an **original** deep learning package, capable of diagnosing diabetic retinal disease (DR) with an accuracy above 90% based on OCT images

INTEL

Deep Learning Software Intern

Shanghai, China

Jan 2022 - Apr 2022

- Debugged one error and merged two utility cases in 2.0 version of Intel **open-source** package, Building Large-Scale AI Applications for Distributed Big Data (**BigDL**); contributions have been merged into BigDL 2.0 formal release
- Cleaned 10 GB raw data of recommendation system with Apache Hadoop and Spark, with which a Deep Interest Evolution Network (**DIEN**) was parallelly trained through Intel Dien, recommending correctly to 86% of all users
- Migrated a PyTorch deep learning example of segmenting brain tumors with **UNet** into BigDL, accelerating the CPU training speed by **20%**

EBAY

Machine Learning Engineer Intern

Shanghai, China

Jan 2021 - Apr 2021

- Published a paper as the **first author** on 31st International Conference on Information and Knowledge Management (CIKM 2022) on the topic of fraudulent transaction detection <https://dl.acm.org/doi/10.1145/3511808.3557136>
- Queried and processed 400GB raw data from eBay online message data warehouse with **SQL** and **PySpark**; performed **HDBSCAN** for extracting data with more specific semantics and removing autoreply text
- Trained a **NLP** model, **DistilBERT** with test data, correctly identifying 92% of fraudulent transactions; packaged above frauds detection pipeline with **Docker** images, and uploaded it to the eBay server for daily online inference

COURSE PROJECTS

LEMONDB: SELF-DESIGNED MULTITHREAD DATABASE

Research Project of VE482, Introduction to Operating Systems

Shanghai, China

Oct 2021 - Nov 2021

- Designed a relational database in C++ that supports basic SQL queries and several database build-in functions
- Improved query speed by 60% through reconstructing a **multithread** version of database with a C multithread library
- Optimized efficiency of multithread database by building a query **scheduler**, rearranging order of read-only queries

MILLION SONG PROJECT

Research Project of VE472, Methods and Tools for Big Data

Shanghai, China

May 2021 - Aug 2021

- Developed a high-speed song recommendation system on Million Song Dataset (MSD) of Columbia University
- Combined 4000 separate .h5 files of MSD into a single avro file with I/O utilities implemented with Apache Avro
- Executed **Parallelized Breadth First Search** in MSD through Apache Spark and Apache Hadoop, and increased speed of data processing (**MapReduce** - 2 min 30 s, **Spark** - 43 s on a 10 GB Subset)

SKILLS

Programming: Python, PyTorch, TensorFlow, Pandas, Hadoop, PySpark/Spark, SQL, C/C++, Docker, Shell Scripting