

Liangwei Zhang

(412) 996-7418 | bbxzpl@gmail.com | [Linkedin](#) | [Personal Website](#)

Education

Carnegie Mellon University – School of Computer Science **Pittsburgh, PA**
Master of Science in Intelligent Information Systems (GPA: 3.89/4.0) Aug 2021 – Dec 2022

Selected Coursework: Machine Learning, Deep Learning, Machine Learning in Production, Search Engines, Advanced NLP, Introduction to Computer Systems

The Hong Kong University of Science and Technology (HKUST) **Hong Kong**
Bachelor of Science in Data Science and Technology (GPA: 3.8/4.3) Sep 2017 – June 2021

Minor in Big Data Technology First Class Honor Degree Classification

Coursework: Algorithms and Data Structures, Cloud Computing, Big Data Mining, Object-Oriented Programming

Skills

Programming Languages: Python, C/C++, Java, R, Javascript, SQL, MATLAB, HTML/CSS, Presto, Haskell

Tools & Libraries: Pandas, Pytorch, Numpy, Tensorflow, Spark, Scikit-learn, Hadoop, Kubernetes, React.js, Linux

Professional Experience

Meta Platforms, Inc. – Software Engineering Intern | Python, Presto, Haskell, Hack May 2022 – Aug 2022

- Developed user models to identify bad actors & victims in the Friending Ecosystem.
- Performed advanced offline data analysis to identify performance gaps in the predictive performance of ML workflows and ameliorate data segments where it underperforms.
- Ran large scale A/B tests to validate the performance of the user models and demonstrated a reduction of X million in bad friend requests.

Tencent Holdings Ltd. – Machine Learning Intern | Python, Spark Feb 2020 – Aug 2020

- Co-developed and improved a 4k-line News Article Recommendation BERT Framework with Python and Pytorch for Tencent News Client Platform used by 100,000+ users.
- Collected and pre-processed over 200,000 articles using Spark and SQL, and optimized the performance of extracting similarities between articles through embedding vectors by performing stopwords filtering through Pytorch, which increased the accuracy of TFIDF by 10%
- Developed Heatmaps by Python to visualize the result from generating similar articles

Yillion Bank – Software Engineering Intern | Python Jun 2019 – Aug 2019

- Developed a Risk Management Framework to estimate probabilities of customer loss default on loans based on personal information
- Developed various analytical models through Python, e.g. RandomForest, XGBoost, and Gradient Boosting Decision Tree (GBDT)
- Improved performance of all models by over 15% and obtained the best performance around 75%

Projects

Image-Based Face Classification CMU | Feb 2022 – Apr 2022

- Implemented one recent baseline model ConvNet toward the design of a vision Transformer
- Applied data augmentation techniques to images including random augmentation, random affine
- Achieved over 90% on test dataset and Ranked 14/354 on Kaggle Competition

Paper to Poster – Directed Study CMU | Jan 2022 – Apr 2022

- Applied current layout generation models to generate layouts of posters based on academic papers
- Devised bounding box annotation instructions of academic papers and annotated over 500 posters

Search Engines Architecture CMU | Sep 2021 – Dec 2021

- Developed software architecture components including the design and implementation of large-scale, distributed search engines over 588,000 documents in dataset.
- Committed 3k+ lines of Java code to implement exact-match and best-match retrieval algorithms.
- Applied SVM ranking framework and diversification approaches to refine the quality of retrieved queries and improve the performance of the search engine by 30%

COVID-19 Knowledge Graph Construction - Final Year Project HKUST | June 2020 – May 2021

- Constructed a comprehensive knowledge graph framework of COVID-19 related information from medical articles
- Designed entity extraction and relation extraction pipelines through pre-trained NER model and OpenIE
- Committed 4k+ lines of Python code to retrieve over 21,000 entities and relations with the accuracy of over 85% in the evaluation and stored them in MongoDB by Graphic User interface
- Built a frontend web application including Search Engines and Graphin API by React and Node.js to visualize the backend data to make over 30 clusters