GitHub: Junze-Zhang
Linkedin: Junze Zhang
https://junzezhang.com

Junze Zhang

(343)-843-3661
zhang_junze@outlook.com
Waterloo, ON (willing to relocate)

Education

University of Waterloo — **Waterloo, ON, Canada** — **Sep. 2023 – Jan. 2025 (Expected)**
- **M.Eng. in Electrical and Computer Engineering (Software Specialization)**
- Average Grade: 91/100

Beijing Institute of Technology — **Beijing, China** — **Sep. 2019 – Jun. 2023**
- **B.Eng. in Data Science**
- Average Grade: 88.4/100 (Top 20%)

University of Cambridge — **Online Summer School** — **Sep. 2022 – Aug. 2022**
- Topic: Deep Learning and Neural Networks
- Instructor: Prof. Pietro Lio

Employment

Software Developer (Intern) — **Beijing Ligare Protocol (Startup)** — **Apr. 2023 – Aug. 2023**
Beijing, China
- Used **Spring Boot** Framework to develop the back-end of a shopping application for shoppers/distributors, and a shop management panel for sellers
- Provided **OpenAPI** (previously Swagger) specification for the Spring Boot framework
- Used **Git** to standardize workflow and software versioning on previously unorganized code
- Identified software defects, provided urgent bug fixes and incremental optimizations, and updated the source code from **Java 8/Spring Boot 2** to **Java 17/Spring Boot 3**
- Collaborated with the Marketing department, identified market needs, and analyzed feasibility of major requirements.

Projects

InsightLabel, A Powerful Image Annotation Tool
- Used **Django Framework** to develop the back-end of annotation data management module and user interface
- Provided **OpenAPI** (previously Swagger) specification for the Django framework
- Utilized **AWS EC2** for system deployment, **AWS RDS (MySQL)** as cloud relational database service, and **AWS S3** for cloud object storage
- Utilized a **Mask R-CNN neural network** to provide automated annotation for images, reducing human work by $\sim$80%
- Designed a comprehensive testing suite with 100% statement coverage, and identified and provided fixes for the software system
- Collaborated with team members using issue boards, with threads for progress, new features, and bugfixes

Research on Pretrained-Model-Driven Methods for Technology Article Explanation [Bachelor Graduate Project]
- Given technology article as context, conducted an experiment on multiple Large Language Models (LLMs) to evaluate their question answering capability
- Given long technology articles and questions, utilized **Dense Passage Retrieval (DPR)** to extract evidences to the question, reducing prompt length by 80-95%
- Proposed a more optimized data preprocessing method for DPR evidence retrieval which increases evidence retrieval accuracy by $\sim$4%

Hate Speech Detection of Meme Pictures Based on Multimodal Machine Learning [Summer School Project]
- Implemented **MMBT** multimodal neural framework, **CLIP** image encoder, and **HuggingFace BERT** language encoder to develop a multimodal neural network; the neural network takes both meme images and attached texts as input, detecting hate speech from combined input
- Achieved an AUROC score of 0.81 with the multimodal network, a $\sim$4.5% improvement over the state-of-the-art
- Developed a web application with **Flask** to use the model; included OCR to extract text from meme pictures

Encrypted Traffic Classification with Multimodal Machine Learning

- Researched over existing encrypted traffic classification problems and models, designed a 3-part CNN-based multimodal neural network, and achieved 99% accuracy on the CESNET-QUIC22 dataset

Movie Recommender System with Apache Spark

- Implemented the Matrix Factorization algorithm with **Apache Spark** data processing framework, with a Root Mean Squared Error of 0.81 on MovieLens dataset

MKBQA – Knowledge Graph Based Question Answering for Medical Inquiries

- Crawled medical data from XunYiWenYao (a Chinese disease database website), extracted the relationship, and built a knowledge graph with **Neo4j Graph Database**
- Utilized a BERT-based model to classify user intent and built a chatbot with Flask

Facebook Social Network Data Analysis and Visualisation

- Used **NetworkX** python library to analyze the statistics of Facebook users, and **Gephi** software to visualise the social network
- Utilized edge bundling algorithm to simplify and beautify the visualisation graph

ML-based DeepFake Identity Theft Prevention

- Analyzed the technological details and ethical issues of DeepFake technology
- Used **EfficientNet** neural network to train on the DeepFake-TIMIT dataset, and developed a tool to detect and give warnings of potential DeepFake pictures or videos

Languages and Technologies

- **Languages:** Python, Java, JavaScript, C/C++, C#, Scala, SQL, HTML, Linux Bash, Cypher(Neo4j)
- **Software Development:** Spring Boot, Django(Python), Node.js, OpenAPI, Flask, Qt
- **Machine Learning:** Tensorflow/Keras, PyTorch
- **Cloud Computing:** AWS (EC2, RDS, S3, SES)
- **Data Analysis and Distributed Computing:** MapReduce, Apache Spark, MySQL, Pandas(Python), Neo4j, Matplotlib
- **Version Control:** Git
- Familiar with Linux command line, Linux server setup, and home internet setup