

# Guoyao Hao

☎ 848-391-6356 | ✉ ghao004@ucr.edu | 📱 haoguoyao | 🌐 haoguoyao

## Summary

Highly skilled software engineer with expertise in AI, deep learning, large language models (LLM), and full-stack development. Two Master's degrees in Computer Science and Computer Engineering. Three years of hands-on experience in intelligent system, NLP, data science and cloud service.

Has a proven record of full-cycle software development, including requirements analysis, design, data preprocessing, model development, testing, and deployment. With the strong ability to integrate the latest techniques and research background, solved real-world business problems, implemented complex machine learning models, and delivered impactful results.

## Skills

<b>Programming</b>	Python, C++, SQL, NoSQL, JavaScript, Typescript, HTML, CSS, C, JAVA, Scala
<b>AI</b>	Pytorch, Tensorflow, LLM, Transformer, Generative AI, NLP, LangChain, Vector database
<b>Data science</b>	Pandas, Numpy, Airflow, Seaborn, Matplotlib, SciPy, Sklearn, Spark
<b>Full-stack</b>	React, Django, Flask, Node.js, Kubernetes, Docker, Nginx, Tailwind
<b>AWS</b>	Elastic beanstalk, API Gateway, Opensearch, S3, Sagemaker, DynamoDB, Lambda, EC2
<b>Tools</b>	Github, VScode, Cursor, Jupyter Notebook, Google Colab, Vim, QGIS
<b>Hands-on</b>	RAG applications, Fintech, Full-stack development, Bioinformatics

## Work Experience

### RAG application for semantic search on PDF document

Buena Park, CA

Position: AI Engineer Intern; Data Application Lab

Jan. 2025 - present

- Developed an interactive web application using **Node.js**, **React**, and **Flask** that allows users to upload PDF documents, query them, and receive streaming results.
- Devised a PDF parsing workflow to structure documents and construct metadata, improving retrieval accuracy.
- Leveraged **Llama Index** to create multiple document embeddings, stored in a **Chroma** database.
- Employed **TruLens** to evaluate retrieval performance and iteratively enhance the final output quality.

### Quantitative trading strategy development

Riverside, CA

Position: Software Engineer; Tazzy Street LLC, Startup

Jan. 2024 - Dec. 2024

- Designed **RESTful API** and **WebSockets** to gather real-time trading data from OKX. Used **Pandas** and **NumPy** to data analysis and feature engineering. Design custom **crawlers** for crypto news from various source.
- Implemented and fine-tune a multi-factor quantitative trading strategy on **BackTrader** and conducted back-testing.
- Developed a user interactive website using **Django** and **React**. Visualize the performance of different trading strategies, and search for best algorithm parameters.

### Predict multiple-site mRNA Alternative Splicing

Riverside, CA

Position: Research assistant. University of California, Riverside

Nov. 2021 - June. 2023

- Targeted the weaknesses of the current **CNN** model, took the competition between splicing sites into consideration, implement and fine-tune a transformer-based model for predicting alternative splicing sites usage.
- Combined **interval-aware LSTM** with a **BERT** model for inference on genes with up to 256 splicing sites. On the GM12878 cell line, it improved the Spearman **correlation from 0.72 to 0.78**, surpassing the state-of-the-art method.
- Analysis deciding factors of the splicing behavior and additionally integrate 11 new epigenomic features to the RNA.

### Cross-domain Multi-attribute Conditional Image Generation

New Brunswick, NJ

Position: Research assistant. Advisor: Vladimir Pavlovic, Department of Computer Science, Rutgers

June. 2020 - May. 2021

- Trained a **GAN** that created high-quality pizza images (FID=8.4) from ingredients and geometric view attributes.
- Created a **3D model** for generating a second dataset with view attributes beyond the range of natural images. Extended StyleGAN2 and trained additional classifiers to give the model **cross-domain attribute control ability**.
- Built a web page for user interaction and demonstration using **Flask** and **Streamlit**, published a paper.

## Interactive Data Visualization and Analysis for NYC taxi data

New Brunswick, NJ, Fall 2020

- Developed an interactive web system to visualize 2005 New York City taxi records (about **26 million** records), complemented by **QGIS integration** for enriched geospatial visualization.
- Included a suite of interactive graphs and charts to intuitively extract and analyze information.
- Opted for **MongoDB** as the database and Flask for server. Employed various techniques to ensure response speed.

## Face Detection based Android application for Attendance Check

Liaoning, China, 2019

- Trained a **YOLO** network for face detection on group photos, applied NMS to remove duplications. Automated face recognition on detected faces and individual photo pairs.
- Designed and Implemented an **Android** application, and a Java server for a two-step attendance check process for classes with up to hundreds of people. Utilized **MySQL** and **FastDFS** for data storage.

## Publication

- Fangda Han, Guoyao Hao, Ricardo Guerrero, Vladimir Pavlovic, Multi-attribute Pizza Generator: Cross-domain Attribute Control with Conditional StyleGAN, BMVC 2021
- Fangda Han, Guoyao Hao, Guerrero, R., Pavlovic, V. (2020). MPG: A Multi-ingredient Pizza Image Generator with Conditional StyleGANs. arXiv preprint arXiv:2012.02821.
- Guoyao Hao, Yifei Li, Bone Age Estimation with X-ray Images Based on EfficientNet Pre-training Model 2021 J. Phys.: Conf. Ser. 1827 012082

## Awards

2021 Graduate Fellowship

Riverside, CA

2021 Dean's Distinguished Fellowship

Riverside, CA

## Education

### University of California, Riverside (UCR)

Riverside, CA

M.S. in Computer Engineering (Overall GPA: 3.96/4.0)

Sept. 2021 - Dec. 2023

- Research: Predict mRNA alternative splicing using transformer based model

### Rutgers, The State University of New Jersey

New Brunswick, NJ

M.S. in Data science

Sept. 2019 - Jun. 2021

### Northeastern University

Shenyang, Liaoning, China

B.E. in Software Engineering

Sept. 2015 - Jun. 2019