Chaoran Huang

Providence, RI | 949-522-1904 | chaoranhuang97@gmail.com | linkedin.com/in/chaoran-huang-8388b7203 | github.com/Chaoran-Huang

TECHNICAL SKILLS

Languages : JAVA, Python, C/C++, C++, SQL (MySQL), JavaScript, TypeScript, Assembly
Frameworks :Spring Boot, Flask, Django, Pytorch, Tensorflow, SpaCy, Crypto++, jQuery, React, Angular, Node.js, ExpressJS, Bootstrap
DevOps and API Tools : Git, Docker, Kubernetes, Jekins, Postman, Jira
Cloud and Security Tools : *Amazon Web Services (AWS)* , *Google Cloud Platform (GCP)* , SQL Server, Linux (Configuring and Managing)

EDUCATION

Brown University	MS in Computer Science, GPA: 3.8	Sept 2022 - May 2024
University of California, Irvine	BS in Computer Science, Major GPA: 3.9	Sept 2017 - June 2021

EXPERIENCE

Software Developer | *My Car Auction, Inc, Irvine, CA*   *March 2021 - July 2022*
  – Contributed to the development of an automated car inspection application and management system, including web app and software development and micro-service integration to enhance business operations.
  – **Project 1: Competitor Price Scraping System**
    - Designed and developed a web scraping system using *Selenium and Puppeteer/Pyppeteer* to fetch competitor pricing data
    - Leveraged *Axios/Express, AWS DynamoDB, and Vaadin framework* for data processing and presentation
    - Automated the system with ngrok and researched anti-scraping techniques for efficiency
    - Successfully scaled the system to scrape over **30000** competitors' auction vehicle prices daily across the entire U.S., providing comprehensive market insights.
  – **Project 2: Oracle NetSuite Enterprise Resource Planning (ERP) and Customer Relationship Management (CRM) System**
    - Developed and integrated NetSuite CRM and ERP system software with iTextPDF, DocuSign, PandaDoc, and Auto-lead Data Format (ADF) using *Java Spring Boot, OAuth2.0, and Retrofit2*
    -Coordinated and customized company's Financial & Accounting team's requirements in collaboration with Oracle NetSuite consultants, greatly boosted working efficiency
  – **Project 3: Named Entity Recognizer for Used Vehicles (Github)**
    - Designed *Natural Language Processing model* with *SpaCy* to recognize important attributes from customers' descriptions and facilitate search engine algorithm to retrieve regulated vehicle information
    - Completed a workflow from inital data collection & cleaning, model training & deployment and software integration.
    - Dived into different models: *CNN, LSTM and transformers: BERT* , analyzed their performance based on their strength and weakness.
    - Enabled offshore team to rapidly receive valuable and pertinent information, greatly improved the auto-lead number by **70%** per day.
  – **Project 4: Sales Visualization Dashboard Web App**
    - Developed an interactive dashboard using *Flask, Angular, and ECharts* for visualizing sales data, enabling better marketing decisions and automating commission calculations for employee payroll
    - Integrated micro-services for offshore teams and the Finance Department to track sales processes and vehicle auction stages

RESEARCH PROJECTS

Full-Stack Developer, Co-founder | *Shuxiang Fayun Technology Co., Ltd., Shanghai, China*   *August 2023 – June 2024*
  – Spearheading the development and integration of an all-in-one business management webapp and a QA-focused *large language model (LLM)* with machine learning and deep learning solutions: *Langchain* for legal professionals. Played a pivotal role in designing the system architecture and workflow, ensuring efficient, scalability and seamless operation.
  – **Project: Integrated Issue Management Webapp and Language Model for Legal Practice**
    - Orchestrated the system design, defining the architecture and workflow logic to optimize communication and case management in legal practices.
    - Built the webapp using full-stack technologies, focusing on *React, and Tailwind* for the frontend, and *Spring Boot with AWS DynamoDB, ElasticBeanstalk, EC2* for the backend.
    - Developed the LLM with PyTorch and Transformers to aid lawyers in legal case understanding and document preparation.
    - Integrated advanced fullstack components: *Next.js* , *Google SSO* , *Prisma* and *Spring Boot Retrofit2, Spring Security* for enhanced functionality.
  – **Achievements:**
    - The bespoke system design significantly streamlined communication and workflow within law firms.
    - The project was widely adopted by multiple law agencies and solo practitioners, also attracted a considerable investment.

Activation Checkpointing for Deep Neural Network Training | *Harvard University*   *Jan 2024 - May 2024*
  – Developed and implemented techniques to optimize memory usage in deep neural network (DNN) training using PyTorch.
  – Created a comprehensive *computational graph profiler* to analyze memory and computation time for each operation, categorizing inputs/outputs, and generating peak memory usage graphs.
  – Designed and implemented an algorithm to selectively store and recompute activations during training, reducing peak memory requirements by up to *70-85%.*
  – Developed tools to manage activation storage and recomputation, ensuring efficient gradient computation during the backward pass.
  – Achieved significant memory savings, enabling the training of larger models with larger mini-batch sizes, and provided detailed experimental analysis on performance improvements.

Ads Exchange Simulation Analyst | *Brown University*   *Jan 2023 - May 2023*
  – Excelled in a rigorous Ads Exchange Game project, securing a *top-3* finish among competitors through auction strategy and advanced predictive analytics.
  – Conducted an in-depth analysis of diverse auction models including *First/Second Price Auction* and the *VCG Mechanism* , applying these insights to significantly enhance bidding strategies for increased profitability.
  – Implemented cutting-edge machine learning techniques by developing a *Boltzmann Machine* combined with *Q-learning* . Integrated these with a *Viterbi algorithm-powered Hidden Markov Model* to refine prediction accuracy and system efficiency in real-time bidding environments.

Data & Web App Development: Netflix Fullstack Clone | *University of California, Irvine*   *March 2020 - June 2020*
  – Developed a Netflix clone movie store website using *Java* and *JavaScript* , managing the backend database with *MySQL* , and deploying the project via *Amazon AWS* for an average load time under **300ms in pressure text**.
  – Enhanced website functionality by implementing full-text search, auto-complete, stored procedures, and various performance-tuning techniques.
  – Strengthened website security with ReCAPTCHA and encrypted passwords, while expanding versatility by creating an accompanying *Android* mobile app with *Google firebase* .