

Yuanjun Chai

+1 206-954-3595 | yjchai@uw.edu | www.linkedin.com/in/yjchai | yuanjunchai.github.io

Education

University of Washington, Seattle Master's in Electrical Engineering Computer Science Track GPA: 4.0/4.0	09/2024 – 06/2026
University of Cambridge, UK Exchange Major in Artificial Intelligence GPA: 4.0/4.0	07/2018 – 09/2018
Xidian University, CN Bachelor Major in Electronic and Information Engineering GPA: 3.85/4.0 Rank: TOP 5	09/2015 – 06/2019

Experience

Machine Learning Engineer , VMware – Beijing, CN <ul style="list-style-type: none">Created a new enterprise level RAG system with fine-tuned LLM, multi-way recall and BGE reranker, deployed it from scratch with company PDF documents, codebase, and images on VMware's cloud-native platform. https://huggingface.co/VMware/open-llama-7b-open-instruct, https://github.com/vmware-private-ai.Developed an educational grading LLM system using a handwritten OCR detector, prompt engineering and LoRA tuning, achieving 99% accuracy of Math objective questions and 81% accuracy of Math subjective questions.Built an AI content moderation algorithm to combat LLM hallucination, achieving 89% accuracy in multi-modal detection of sensitive content through fine-grained dataset categorization and comprehensive model training across content moderation classification.	07/2022 – 09/2024
Machine Learning Engineer , Alibaba Group – Beijing, CN <ul style="list-style-type: none">Developed KreadoAI – an AI virtual avatar platform, optimized algorithms such as lip-sync alignment, voice cloning, and background effects of PPT digital human in videos, resulting in \$1M+ quarterly revenue across Europe, Africa, Southeast Asia, and the Americas. https://kreadoai.com.Designed new datasets and new training strategy to improve digital human video generation quality, achieving a 45% enhancement in lip synchronization and image quality, and boosting user subscription rates by 38%.Implemented AI algorithms integrating Stable Diffusion, LoRA, and XL to optimize image generation for product-specific applications (e.g., glasses, wigs), creating thousands of visually appealing marketing posters.	07/2021 – 07/2022
Research Assistant (Full-time) , University of Hongkong – HongKong, CN <ul style="list-style-type: none">Developed the algorithm framework for identifying the spine region and detecting the spine key point.Designed and implemented the MPF-net algorithm in Python, developed a C++ executable framework for Windows, and deployed it to assist doctors, improving clinical diagnosis accuracy by 10% and efficiency by 59%.Collaborated with doctors to annotate medical image datasets using ImageJ, creating spine key points, bounding boxes, and segmentation training sets. Got HKU PhD return offer.	06/2020 – 07/2021
Research Assistant (Full-time) , SIAT – ShenZhen, CN <ul style="list-style-type: none">Design an industrial-grade super-resolution framework for multi-GPU distributed training and testing, achieving state-of-the-art results on the two largest datasets and producing high-definition video frames (4K, 8K).Developed an iteratively corrected blind image super-resolution algorithm for addressing unknown degradation, releasing the project on GitHub with 200+ stars. https://github.com/yuanjunchai/IKC.Designed and maintained the lab website using front-end (HTML, CSS, JavaScript, Bootstrap) and back-end frameworks (Flask, Django), building an image-processing-focused platform that achieved 100k+ page views. http://xpixel.group.	06/2019 – 06/2020

Skills

Programming Languages: C++/C, Java, Python, MATLAB, SQL, CUDA C/C++

Tools: VS Code, Git, PyTorch, TensorFlow, Keras, Linux, Shell, Docker, k8s, Flask, Django, FastAPI, Azure Cloud