Zhexian Zhou

412-918-0956 | jakozhou@gmail.com | linkedin.com/in/zhexianzhou | github.com/JakoError | jakoerror.github.io

Education

Carnegie Mellon University — Pittsburgh, PA
*Master of Science in Artificial Intelligence Engineering (ECE) – GPA: **4.0*** — *Aug 2024 – Dec 2025*

University of Electronic Science and Technology of China (UESTC) — Chengdu, China
*Bachelor of Engineering in Software Engineering (Elite Program) – GPA: **3.94*** — *Aug 2020 – July 2024*

Experience

Research Assistant — Sep 2024 – Present
CMU AirLab & Team Chiron - DARPA Triage Challenge

- Developed vision modules and ROS2–based end-to-end edge inference and benchmark pipeline for multimodal VLMs triage VQA across UAV/UGV platforms, API supporting SOTA VLMs (QwenVL, InternVL, Llama, VILA)
- Proposed RGBT knowledge-transfer and SFT fine-tuned NVILA improved IR Hemo Acc from 36.6% to 54.1%
- Proposed and built quantization + multi-turn KV-cache for AGX Orin, achieving real-time <1s follow-up latency
- Developed Air-Infer project to stream ROS2 Messages and Vision Tensor over gRPC/HTTP for robust inference
- Accepted (*ICLR 2026*) "KnowledgeSmith: Uncovering Knowledge Updating in LLMs with Model Editing and Unlearning." (Luo, Y., **Zhou, Z.**, Chen, H., Qiu, K., Savvides, M., Li, S., Wang, J.)

Undergrad Researcher — Jan 2023 – Jun 2024
UESTC Memristor & Digital Information System Research Lab

- Outstanding Thesis (UESTC 2024). Proposed dynamic confidence and edge-aware semi-supervised learning for sparse segmentation; mIoU improved 70.5→76.0→83.9→86.9 as the supervision rate increased 10→20→50→100
- Proposed geometry-aware skeletonization and route-pattern recognition for tubular structures (e.g., blood vessels)
- Published conference papers on fixed-point FFT and evolutionary algorithms using memristor crossbar circuits

Researcher — Jun 2022 – Jun 2024
UESTC & TibetU Tibetan Language Automatic Recognition Technology

- Built diffusion-based OCR/scene-text augmentation platform integrated into the project training pipeline system
- Selected as the outstanding undergraduate researcher on MoST CN 2030 Key R&D Program (No.2022ZD0116100)

Projects

Adaptive Reasoning for Vision-Language Models — Jan – Apr 2025

- Proposed LoRA fine-tuning with GRPO reward, enabling step-by-step reasoning; increased overall accuracy from 59.81% to 62.01% across image/video datasets (MathVision, VQA-CP, LLaVA-150k, Video-ChatGPT, Shot2Story)

Feature-Fusion Face Detection — Jan – Apr 2025

- Developed HOG/PCA/K-means feature-fusion on SVR/AdaBoost/YOLO, achieved 98.09% mAP50 on validation

MatSAR - C++ Math Toolkits Development — Aug 2022 – Aug 2024

- Architected and developed cross-platform C++ toolkit with Multi-dimensional Matrix Core, Matrix Operations, and Math Functions. Led developer team to develop over 200 algorithms across Linear Algebra, DSP, and Calculus

C++ Database Management System — Oct 2022 – Feb 2023

- Designed and developed server–client DBMS supporting multiple data types and common operations; built core storage and query components; Open-sourced on GitHub: JakoError/cppDBMS

Compiler Toolkits for SysY programming language — Dec 2021 – Jan 2022

- Developed compiler using Flex/Bison, including lexical analyzer, syntax parser, intermediate-code generator, and basic optimizer; Open-sourced on GitHub: JakoError/CompileStudy

DiaryInUESTC - Diary Mobile App — Sep – Dec 2021

- Developed diary, bookkeeping, and memo application with AMap (Gaode) location features and a MyBatis-based persistence layer; Open-sourced on GitHub: JakoError/DiaryInUESTC

Technical Skills

Languages: C/C++, Python, Java, MATLAB, SQL, JavaScript, HTML/CSS
Frameworks: ROS, PyTorch, PyTorch Lightning, TensorFlow/Keras, NumPy, gRPC/REST, PySpark, Spring Boot
Developer Tools: Git, Docker, AWS, VS Code, Visual Studio, GCC, CMake