Yike Tan

http://likegiver.github.io

Email : yiketn@gmail.com

Mobile : +1-412-583-0350

EDUCATION

Carnegie Mellon University Pittsburgh, PA
Master of Science in Artificial Intelligence Engineering - Information Security; GPA: 4.00/4.00 *Dec. 2025 (Expected)*
- **Coursework**: Large Language Models, AI Systems and Tool Chains, Computer Systems

University of International Business and Economics Beijing, China
Bachelor of Engineering in Data Science and Big Data Technology; GPA: 3.70/4.00 *Aug. 2020 – Jul. 2024*
- **Coursework**: Machine Learning, Data Structure and Algorithm, Natural Language Processing, Big Data Analysis

EXPERIENCE

Epoching AI Beijing, China
Machine Learning Engineer Intern *Jul. 2024 - Aug. 2024*
- Developed an end-to-end watermark removal service using PyTorch, fine-tuned pre-trained ViT model through comparative experiments, achieving 30% accuracy improvement over industry standards.
- Optimized training pipeline with synthetic data generation and augmentation techniques, effectively handling semi-transparent watermarks and logos in e-commerce imagery.
- Deployed scalable FastAPI/Docker service with Prometheus monitoring, handling concurrent requests

Ytell Network Technology Co., Ltd. Beijing, China
Machine Learning Engineer Intern *Sep. 2023 - Mar. 2024*
- Developed multimodal knowledge bot by integrating VLMs (MiniGPT-V, LlaVA) with FastAPI backend, enabling natural language and image-based document search
- Built React frontend and integrated RAG system using LangChain and Milvus vector store, developed automated pipeline generating 50K+ QA pairs from documents, improving response accuracy from 75% to 85%
- Designed and implemented automated order processing pipeline using Apache Airflow, FastAPI, and MongoDB, orchestrating ETL workflows to extract and store chat history and order data, processing 1000+ daily requests with 95% accuracy

PROJECTS

Memory Allocator *Oct. 2024*
- Implemented a high-performance dynamic memory allocator in C from scratch, including malloc, free, realloc, and calloc functions with throughput optimization using segregated free lists and boundary tag coalescing.
- Implemented heap debugging tools with GDB integration for validating block alignment and free list consistency, achieving 75%+ memory utilization.

Transformer Decoder *Sep. 2024*
- Built a decoder-only Transformer from scratch using PyTorch with masked multi-head attention and positional encoding, engineered custom BPE tokenizer and training pipeline with C4 dataset.
- Published trained model to HuggingFace Hub, integrated Weights & Biases for experiment tracking, and developed text generation pipeline with temperature-based sampling

Self-LLM: Open-Source LLMs Deployment Guide (10+k Stars on Github) *May 2024 - Jun 2024*
- Collaborated with team members to create comprehensive deployment guides for GLM4, including LangChain integration, FastAPI, WebDemo, vLLM deployment and LoRA fine-tuning, enabling developers to build advanced LLM applications on Linux platforms.

Keep Knowledge in Perception: Zero-shot Image Aesthetic Assessment *ICASSP 2024 (Oral)*
- Developed a novel zero-shot framework using CLIP for fine-grained aesthetic assessment across 7 attributes (composition, color, lighting), outperforming SOTA approaches.
- Created a large-scale aesthetic dataset with 175K expert critiques, enabling knowledge transfer from photography expertise to model perception.

PROGRAMMING SKILLS

- **Programming Languages:** Python, C, C++, Java, JavaScript, SQL, R
- **Packages:** React, Node.js, Django, Flask, NumPy, Pandas, PyTorch, TensorFlow, Jax, SparkML
- **Tools:** Git, Docker, Kafka, Neo4j, Google Cloud Platform, Unix, LaTeX