Hui Xu

(516) 457-4066 huixucom@gmail.com linkedin.com/in/huihsuxu github.com/huixu11 huixu11.github.io

EDUCATION

State University of New York at Stony Brook **Stony Brook, NY**
MS in Engineering Artificial Intelligence — GPA: 3.67/4.0 *Aug 2024 – Dec 2026*

- **Coursework:** Distributed Systems, NLP, Deep Learning Algorithms, Robotics, Statistical Learning
- **Publication:** MAQuA: Adaptive Mental Health Screening using IRT (EACL 2026) [arXiv]

Beijing Forestry University **Beijing, China**
MS in Computer Application Technology (4.0/4.0) — BS in Information Systems (3.9/4.0) *2009 – 2018*

TECHNICAL SKILLS

Languages: Python, TypeScript/JavaScript, C/C++, Java, SQL **Infra:** Docker, Git, AWS/GCP, Ray; Multi-Paxos, 2PC
ML/LLM: PyTorch, TensorFlow, LangChain **Backend:** Django, Flask, FastAPI, PostgreSQL, SQLite, Redis, MySQL

PROFESSIONAL EXPERIENCE

Mastercard **Beijing, China**
Software Engineer II *Nov 2021 – Jul 2024*

- Developed **full-stack business analytics platform** ("Test & Learn") for Chinese banks using **Django (backend), React/Redux (frontend), and PostgreSQL/SQLite**, enabling local deployment and data compliance
- Proposed and prototyped migration from in-house multiprocessing framework to **Ray Core**, enabling **distributed execution across clusters** and containerized environments with minimal code changes
- Engineered high-performance **outlier detection algorithm** using statsmodels leave-one-out statistics; **reduced runtime from 9 hours to 10 minutes (98.1% improvement)** through selective computation and vectorization
- Built **Metric Uploader** feature capable of processing **400MB+ CSV files in under 60 seconds** with row-level validation using fully vectorized algorithms and comprehensive unit testing coverage
- Designed and delivered Driver Summary module featuring driver significance analysis and visual summaries using React, Redux hashmaps, and Recharts; collaborated with Product Managers and Tech Leads to refine requirements

Dazhangfang (Chinese Intuit) **Beijing, China**
Python Engineer *Jul 2018 – Oct 2021*

- Managed and deployed **large-scale OCR platform (100,000+ lines of code, 10 servers)** integrating recognition engine, invoice verification service, and web service for automated receipt recognition and accounting
- **Optimized database queries and indexing**, improving Invoice Recognition Web Service performance by **99.99%**, dramatically reducing response latency
- Automated invoice verification process, achieving **90% reduction in human intervention** using edit-distance algorithms for text matching and validation
- Implemented **asynchronous task scheduling and message delivery** using APScheduler and Redis as message queue broker, increasing throughput and system reliability

PROJECTS

2026 Berkeley AgentX–AgentBeats competition – Finance Agent (Sponsored by OpenAI 2nd place) | *Team lead*

- Built a multi-agent financial analysis benchmark system using the A2A protocol, achieving **2nd place** in the Berkeley AgentBeats Competition, covering 18 evaluation categories, 6 MCP data servers, and adversarial debate mechanisms.
- It combines expert-verified static tasks with adversarial trading simulations scored purely on realized performance.

Scalable Distributed Transaction System | *DistAlgo, SQLite, Multi-Paxos, 2PC*

- Designed and implemented **9-node, 3-cluster replicated transaction system** with Multi-Paxos for intra-shard consensus and **Two-Phase Commit (2PC)** protocol for cross-shard atomic transactions
- Achieved **sub-1.5s leader failover** through heartbeat-based failure detection and automatic re-election; implemented configurable cluster topology (N clusters × M nodes) with automatic shard map generation

Transformer Language Model from Scratch | *Python, PyTorch, NLP, Transformers*

- Engineered Transformer LM from scratch with **BPE tokenizer, RMSNorm, RoPE, multi-head attention, and SwiGLU** layers; built and tested Linear, Embedding, and Attention modules ensuring stable gradients
- Trained on TinyStories/OpenWebText using multiprocessing pre-tokenization and custom **AdamW optimizer**; achieved sub-2 minute BPE training (10K vocab) with fluent text generation and competitive perplexity

SERVICE

Reviewer: Agents in the Wild: Safety, Security, and Beyond (ICLR 2026 Workshop)

ACHIEVEMENTS

Agentic AI MOOC legendary tier: Finance agent (sponsored by OpenAI) 2nd place - 2026 AgentX-AgentBeats
Jane Street GPU Mode Hackathon: 10th Place - Optimized latency and accuracy using dynamic batching strategies
2026 Hack-Nation Hackathon: ThoughtLink 1st Place - decoding high-level human intent from non-invasive brain signals and mapping it directly to executable robot instructions.