Zuo Wang Irvine, CA | https://zuo.vercel.app | amangoworks@gmail.com | +1-757-633-7384

Staff Software Engineer. AI Infrastructure & Backend Systems

At Samaya AI, I built and scaled the backend for a financial research agent, serving 10k daily active users from Morgan Stanley, BlackRock, Citi, and Point72. I cut TTFT from 100 to 10 seconds and took a million off the annual infra bill. Previously, I led a seven-person team at Tesla.

WORK EXPERIENCE

Samaya AI London

Member of Technical Staff Jan 2024 - Current

- Architect Knowledge Ingestion Pipeline, Embedding Chunking strategy, Eval, RAG, and AI Infrastructure
- 0-1 turn ML ideas into product features like QueryDecomposition, DeepResearch, and EarningsReport
- Cut latency from Redis, Python, and vLLM
- Demo for our $43.5M series A

Tesla Shanghai GigaFactory

High Voltage Manufacture Org Texas GigaFactory

Staff Software Engineer (manage a team of 7 people) 2022 – 2024

- Responsible for high voltage system for Tesla Model 3, Y, S, X, Powerwall, Megapack, and SuperCharger
- Iterate chip design for Cybertruck and Megapack under tight deadline (2.5 months) and heavy pressure
- Close collaboration with worldwide vendors (CATL, BYD, TI, COSMX, HYC, Pegatron, Jabil). Often travel to their countries to conduct design reviews, on-site inspections, troubleshoot issues, and reduce defect rates
- Teach teams in Manufacture, Process, and Quality to enable them to do their own debugging and fault analysis. This helped making China battery production line 3x faster than USA line
- Finish Cybertruck battery charger 4 weeks ahead of schedule
- Bring up 50 new manufacture stations, maintain over 200 stations, handle $500k budget per year

Atlassian Mountain View, CA

Cloud Security + Platform Infrastructure Fully remote

Software Engineer II 2018 – 2022

- On call and answer questions for a 99.99% availability internal PaaS hosting over 1400 of Atlassian services
- Develop auth sidecar, secure for all communication between Atlassian microservices and staff
- Rewrite tools like ldap cache, rate limiter, and deduper from Java, TypeScript, and Python to Golang
- Create instant generator for services with bells and whistles like k8s, SOX compliant, and CI/CD
- Tools: Splunk, SignalFX, Jira, Confluence, AWS S3, DynamoDB, Lambda, Postgres, Spinnaker

Ontra AI Fully Remote

AI architect 2026

- 0 to 1 the entire life cycle of injecting a new agentic AI product line for the company, from exploring ideas, to vibe coding prototype, to MVP, to PMF, to co-development with lawyers, to marketing and sales, to scaling.
- Worked directly with lawyers as the only engineer on the team to build legal AI product for private equity.
- Built the evaluation set, labeled training data, and worked on reward design and supervised fine-tuning.

TECH STACK

Go, Python, TypeScript, Ruby, C++, Java, Rust

aws, gcp, k8s, gRPC, graphQL, vLLM, Arize, TensorRT, VertexAI

EDUCATION

Johns Hopkins University *Master of Science GPA: 3.9/4.0* *2018*
Machine Learning Security*: Cloud Security, Software Vulnerability Analysis, Computer Forensics, Network security, Privacy Laws, Risk Management, Ethical hacking, Modern Cryptography*
University of Rochester *Bachelor of Science GPA: 3.6/4.0* *2014*
Joseph C Wilson Scholarship; Dean's list
ECE*: signal processing, robot control, semiconductor, microprocessor, C++, Java, assembly, VHDL, MATLAB.*
Minor: Mathematics and Economy

PROJECTS

www.agemo.ai/codewords home
- Fullstack webapp for automated website building with NodeJS and Shadcn.

Unknot.id Baltimore, MD
- Implement black-box adversarial attacks against a behavior biometric auth mechanism on Android
- Harden the LSTM model robust against this <1000 queries shadow model attack building upon CleverHans

Docker privilege escalation *2020*
- Dirty COW race condition to escape from container to host

IoT Exploit *2019*
- Discover network vulnerabilities from DJI drones, Bebop drones, and Amazon Echo

Malicious Website Detection through URL Analytics *2019*
- Gather dataset from PhishTank for malicious URLs and DMOZ for benign URLs
- Utilize machine learning models in JupyterNotebook and use ROC curve to measure performances

Drone System Rochester, NY
- Create a drone drop system for National Training Center and Joint Readiness Training Center of U.S. Army
- 3D CAD Design and print remote controller, release mechanism, and program microprocessors in Arduino

Entrepreneurship
- Cheap long distance hitchhiking platform https://prezi.com/view/bhBhNzij1gost2k0CTnM/ *2017*
- Language exchange to help people find English/Chinese speaking partner echinglish.weebly.com *2014*

Github: github.com/MrMegaMango