

Zuo Wang

amangoworks@gmail.com || +1-757-633-7384

Staff level AI backend infra engineer. 6 years of professional experience. \$10M business impact.

Cross-disciplinary technical breadth in Software and Hardware Engineering, ML, Security, Product, and Infra

Experienced people leader, fun to work with, appreciate smart engineers with low ego. Friendly, honest, and genuine

Conducted dozens of algorithm and system design interviews, grew team to 7 people

Proven record in strategize and deliver reliable systems at scale through cross-team collaborations

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
- Decrease time-to-first-token from 100 to 10 seconds (-90%), scale infra from 10 to 10k daily active user
- Onboard customers from MorganStanley, BlackRock, Citi, and Point72. Strategize pricing
- 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)

July 2022 – Jan 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

2018 – July 2022

- Develop auth sidecar, secure for all communication between Atlassian microservices and staff
- Rewrite tools like Idap cache, rate limiter, and deduper from Java, Typescript, and Python to Golang
- On call and answer questions for a 99.99% availability internal PaaS hosting over 1400 of Atlassian services
- 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

TECH STACK

Golang, Python, GCP, AWS, k8s, gRPC, GraphQL, vLLM, Grafana, Arize, MongoDB, VertexAI, Pinecone, Claude

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

SIDE PROJECTS

www.ontra.ai

- Replace thin gpt wrapper Ruby on Rails software to proper ML service that does retrieval, cross-encoding, prompt evaluation. Reduce thousands of hours for our human in the loop lawyers wasted on poor AI output

www.agemo.ai/codewords

home

- Fullstack webapp for automated website building with NodeJS and Shadcn.

Rippling

remote

- Build out global expansion framework for Rippling's payroll product using PythonDjango+MongoDB+React

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

Cognitive Modeling of User Security

Baltimore, MD

Researcher at JHU Computer Science Department

- Build computational cognitive modeling to simulate human decision-making in Python
- Adapt Instance-Based Learning Theory and ACT-R memory chunk activation and analyze 40GB data

Docker privilege escalation

2020

- Research and apply 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

CryptoDoneRight

2019

- Assist in building cryptodoneright.org (a training tool for server hardening)

Leetcode

- Solved 1000+ questions, improve contest rating from 1460 to 2109 in 7 months. leetcode.com/mangoman/

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/bhBhNzj1gost2kOCTnM/> 2017
- Digital library for mixing drinks app 2016
- Language exchange to help people find English/Chinese speaking partner echinglish.weebly.com 2014

OTHER ACTIVITIES

Indie game dev

<https://battle-map-344101.wl.r.appspot.com/>

DIY electronics

headphone amplifier; windshield wiper controller; pc builds

Music

guitar player; UofR carillon society carillonneur. Multiple live concert performance

Sports

standup paddle board enthusiast, top 0.3% LoL player, amateur boxer, 10 pull ups

Personal website: <https://zuo.vercel.app/>

LinkedIn: www.linkedin.com/in/mrmightymango/

Github: github.com/MrMegaMango

Computational Cognitive Modeling research: behavior.isi.jhu.edu/people.html

Paper: <https://www.unknot.id/post/yes-you-can-spoof-behavioral-biometrics-with-adversarial-learning>