Zinuo Li

Abu Dhabi | +8617749742466 | zinuo93@outlook.com | https://zinuo.me | Linkedin | Github

Summary

AI Engineer with a full-stack foundation, dedicated to productionizing machine learning models through robust **MLOps** and **optimized inference**. Proven ability to engineer fast, stable, and monitored AI services, leveraging over 5 years of experience in **Go** and **Java** and recent deep expertise in **Python**. Passionate about bridging the gap between model development and scalable, high-availability product deployment.

Experience

Single Family Office
Principle AI Engineer

Jan 2024 - Present
Abu Dhabi

- **End-to-End Product Leadership:** Led the complete development lifecycle for two AI projects, transforming them from proof-of-concept demos into fully operational products by defining architecture, managing infrastructure, and overseeing deployment.

- **Full-Stack & Cloud Architecture:** Architected and built the core application using Python and React, while managing cloud-native infrastructure on AWS and Azure with Kubernetes to ensure scalability and high availability.

- **MLOps & Automation:** Engineered a comprehensive MLOps framework to automate the model lifecycle, from data ingestion and training to deployment and monitoring, drastically reducing time-to-market for new features.

- **Performance Optimization:** Focused on production readiness by optimizing model inference, leading to faster response times and lower computational costs for real-time AI services.

- **Advanced AI Implementation:** Developed sophisticated AI solutions by implementing autonomous agent frameworks and a Retrieval-Augmented Generation (RAG) system, enabling the product to provide contextually aware and evidence-backed answers.

- **AI Safety & Trust:** Integrated critical content control mechanisms for text moderation and response verification, significantly improving the trustworthiness and safety of the AI's output.

- **Data Engineering:** Built resilient data pipelines using Apache Spark to process high-volume news and market ticker data, providing a clean, reliable data source for the RAG knowledge base and other downstream models.

Autonomic
AIOps Site Reliability Engineer

Aug 2022 - Aug 2023
Shanghai

https://autonomic.com/

- Pioneered an AIOps initiative to automate and intelligently scale SRE operations.

- Developed an intelligent alerting system to auto-triage incidents and recommend runbooks, reducing Mean Time to Resolution (MTTR) by 60%.

- Built a auto respondent platform by collecting data and analyzing traffic patterns to proactively manage surges and prevent system overload.

- Deployed an AI chatbot to simplify Infrastructure-as-Code (IaC) management and enable developer self-service.

Informatica
Full stack engineer

July 2017 - Aug 2022
Chicago, US

https://www.informatica.com

- Played a key role in the company's cloud modernization initiative, contributing to the migration of core services to a scalable, containerized Kubernetes environment.

- Engineered and maintained features for business-critical microservices, including the scheduling engine, customer licensing, and Role-Based Access Control (RBAC).

- Developed a robust file service API responsible for the secure upload, storage, and retrieval of user and system files.

Education

Illinois Institue of Technology
Computer Science

Sep 2015 - Dec 2017
Masters