Ujwal Kandi

📞 (737) 326-8598 📍 Austin, Texas ✉ ujwal.kandi@utexas.edu 💼 linkedin.com/in/ujwalkandi 🌐 www.ujwalkandi.in

Professional Summary

I've kept systems running across 16+ production environments for Disney, Bank of America, Marriott, and now Apple Inc.
Site Reliability (SRE) & Observability Engineer with 3+ years building systems that reduced Mean Time to Resolution (MTTR) by 75%, designed AI-powered edge diagnostics deployed using Docker, and architected data pipelines for Fortune 500 clients. Currently at Apple; M.S. in IT Management from UT Austin.

Core strengths: Incident Response & Root Cause Analysis, Observability (Grafana, Dynatrace, ELK), AWS & Azure, Infrastructure as Code, and AI/ML for Operations.

Experience

Apple Inc. — Jan 2025 − Present
Metadata Operations Engineer (Contract via Welo Data) — *Austin, TX*

- Optimized mission-critical metadata ingestion pipelines ensuring high availability and under 2s latency for Apple TV's global content catalog using automated validation, error handling, and rollback procedures across **XML** workflows
- Established system health dashboards and alerting for metadata pipeline performance to detect and triage data quality issues, reducing feed failures by 30% and improving platform stability
- Implemented structured debugging and pattern recognition frameworks to identify root causes of batch job failures, reducing resolution time by 25% through systematic Root Cause Analysis (RCA) and runbook development
- Designed failover procedures for critical metadata systems, ensuring continuity and compliance with Apple's service-level requirements

Sports Excitement — Aug 2024 − Jan 2025
Data Operations Engineer — *Austin, TX*

- Built and maintained production Extract, Transform, Load (ETL) pipelines using Apache Airflow (Directed Acyclic Graph orchestration, performance tracking, and failure recovery), supporting 50+ daily batch jobs with 99.8% reliability
- Optimized Azure Data Lake Storage architecture, reducing query latency to 550ms and improving downstream ML analytics throughput
- Automated incident notifications via Airflow trigger jobs & alerting workflows, achieving response times within service-level targets
- Implemented data quality checks and load-testing frameworks, reducing post-deployment failures by 28%

Dover Fueling Solutions — Jun 2024 − Aug 2024
Site Reliability Engineer — *Austin, TX*

- Enhanced visibility by implementing 3-pillar observability strategy: configured Grafana dashboards, Azure API Management (APIM) log analytics integration, and Kafka metrics exporters; reduced issue detection time from 15min to 3min and resolution time by 40%
- Designed and deployed agentic AI system for edge device diagnostics on resource-constrained hardware (SQLite + Ollama), enabling field technicians to resolve issues autonomously; reduced support tickets by 35% and improving critical fault resolution to under 10min
- Utilized LangChain to prototype multi-agent configurations and employed LangSmith to evaluate agent performance, response accuracy, and token efficiency in edge inference scenarios while maintaining diagnostic accuracy under 2GB memory constraints
- Implemented alert tuning and intelligent escalation workflows to reduce alert fatigue by 45% while maintaining high critical incident detection rate; documented best practices in runbooks for on-call rotations

Stronghold Investment Management — Jan 2024 − May 2024
AI Engineer - Capstone Project — *Austin, TX*

- Designed and implemented a custom-built AI agent over Azure AI Studio, integrating OCR and LLM (Mixtral) models to enhance ownership verification processes and automate complex document workflows
- Reduced document processing costs by 20% by implementing an automated document intelligence system for document analysis
- Conducted model response evaluations across multiple Azure-hosted models using custom test datasets and task-specific tuning to benchmark performance (accuracy, coherence, latency) across query types
- Structured prompt design evaluations within Azure AI Studio to refine templates and optimize response quality with prompt tuning for document analysis workflows

Epsilon — Apr 2022 − Jul 2023
Production Support Engineer — *Bengaluru, India*

- Orchestrated incident response lifecycle (detection → triage → resolution → RCA → prevention) across 16+ client systems, achieving 99.8% SLA compliance and reducing resolution time from 45min to 12min through proactive system visibility and automated remediation
- Designed and maintained comprehensive observability infrastructure: integrated CloudWatch, Kibana, and Dynatrace; established intelligent baselines, thresholds, and outlier detection algorithms preventing 6 out of 10 potential incidents before customer impact
- Authored 20+ runbooks, SOPs, and postmortem documentation implementing blameless RCA culture; broke repeat incident cycles by 50% through systematic root cause prevention and knowledge transfer
- Implemented Infrastructure as Code business continuity framework: automated failover scripts for data pipelines using CloudFormation, reducing Recovery Time Objective (RTO) from 2hrs to 30min and maintaining high uptime during critical incidents
- Optimized Matillion ETL job orchestration (SLA schedules, retry logic, maintenance windows), improving batch completion efficiency by 32% and reducing operational overhead by 38%
- Resolved critical data load failures within Epsilon's data mart by optimizing complex batch processes, improving system performance by 89% and ensuring consistent data availability across daily loads

Epsilon Data Management, LLC. — Aug 2021 − Apr 2022
Data Programmer I — *Bengaluru, India*

- Engineered and optimized ETL pipelines for 16+ global clients (Unilever, Marriott, Bank of America, Disney, National Geographic), improving data load reliability and reducing batch failure frequency by 45%
- Designed test pipelines and production-like staging environments to simulate high-volume data flows, reducing post-deployment issues by 33% and improving validation accuracy through systematic performance testing
- Automated 50+ batch jobs and implemented CloudWatch/Dynatrace/Kibana observability stack, enabling early issue detection and accelerating root cause identification by 30%
- Resolved critical data ingestion and quality issues across AWS S3 and Redshift environments, achieving consistent data availability for mission-critical analytics platforms

Isthriwala **Jun 2020 – Jul 2021**
Data Management Analyst *Vijayawada, India*
- Optimized database operations for 35,000+ individuals and 50+ B2B clients using SQL, improving query performance
- Developed KPI dashboards in Tableau for real-time business metrics, improving decision-making velocity and enabling 15% improvement in monthly performance targets

Technical Skills

Languages & Scripting: Python, Bash, SQL, YAML, XML, JavaScript, R

Observability & Monitoring: Grafana, Prometheus, CloudWatch, Kibana, Dynatrace, Azure Insights, ELK Stack

Infrastructure & Deployment: Kubernetes, Docker, Terraform, CloudFormation, Ansible

Data Pipelines & CI/CD: Apache Airflow, Kafka, Jenkins, GitLab CI/CD

Cloud Platforms: AWS (EC2, S3, Lambda, Redshift, CloudWatch), Azure (ADLS, Log Analytics, Azure AI Studio, APIM)

AI/ML & LLM: LangChain, Ollama, LangSmith, ChromaDB, Databricks

Tools & Platforms: Git, GitHub, Jira, ServiceNow, Apache Spark, PySpark, Tableau, Power BI

Education

The University of Texas at Austin **Jun 2023 – May 2024**
M.S. in Information Technology & Management *Austin, TX*
McCombs School of Business
Coursework: Distributed Systems, Data Management, Unstructured Data Analytics, Data Governance & Responsible AI

Projects

ARIA – Adaptive Retail Intelligence Agent (First Place, ElevenLabs Worldwide Hackathon)
- Built a real-time, voice-based conversational retail agent using a multi-agent ReAct architecture to detect emotional and behavioral triggers and deliver personalized product recommendations during live customer interactions
- Integrated Groq LLMs, ElevenLabs TTS, WebSocket audio streaming, dynamic customer profile learning, backend on Supabase , and automated post-purchase email workflows (n8n), deployed via Docker for live demo use

FlipFinder Pro – Real-Time AI System for Used-Car Arbitrage
- Built an ML-driven pricing and anomaly-detection system that analyzes used-car market data to identify undervalued listings and high-profit arbitrage opportunities in real time
- Achieved 89%+ price-prediction accuracy and surfaced $20M+ in potential flips through an interactive analytics dashboard with profit, risk, and confidence scoring (Databricks Free Edition Hackathon)

A.I.D.E. – AI-Powered Observability & Root Cause Analysis Engine
- Built an end-to-end observability and RCA system combining metrics analysis, log anomaly detection, and multi-agent LLM workflows (LangChain) for autonomous API troubleshooting.
- Enabled low-latency, on-device diagnostics using RAG (ChromaDB) and Llama Stack with Ollama + Groq, achieving sub-second inference under 2GB memory constraints (Meta ATX Llama Hackathon).

Leadership & Achievements

ElevenLabs Worldwide Hackathon – First Place Winner & Best Product Built with n8n **Dec 2025**
Global Hackathon *Austin, TX*
- Led development of ARIA, a real-time voice-based conversational retail agent using multi-agent LLM orchestration, ElevenLabs TTS, and n8n-powered automation for personalized recommendations and post-purchase email workflows.

Seeed Studio Ranger Program **Oct 2024 – Present**
AIoT Developer Program *Austin, TX*
- Designed and delivered hands-on AI/IoT technical training workshops reaching 100+ university students, building practical skills in artificial intelligence and edge computing systems.

Dover Fueling Solutions Hackathon'24 – First Place Winner **Feb 2024**
Annual Hackathon *Austin, TX*
- Won company hackathon designing conversational AI system with edge computing for fuel station operations; presented solution at DFS User Conference 2024 to executive leadership and industry stakeholders.

MSITM Campus Ambassador – McCombs School of Business **Aug 2023 – May 2024**
Graduate Program Ambassador *UT Austin*
- Guided prospective students through the application process, increasing applications from underrepresented backgrounds by 60%.

University Innovation Fellows **Jan 2018 – Dec 2020**
International Fellowship Program *Stanford d.school*
- Conducted 15+ workshops on Design Thinking and problem-solving reaching 1000+ students across multiple universities; earned scholarship to attend Silicon Valley Meet for immersive learning experiences at Stanford, Google, and Microsoft.