Siddharth Rohit Jain

Tempe, AZ | (623) 326-7382 | tellsiddh@gmail.com | tellsiddh.com | linkedin.com/in/tellsiddh | github.com/tellsiddh

SUMMARY

Experienced software engineer with proven success in developing and deploying scalable software solutions, specializing in ML/AI applications, cloud-native AWS infrastructure, CI/CD, observability, and infrastructure management.

EDUCATION

Arizona State University — Tempe, AZ
Master of Science, Robotics and Autonomous Systems - Thesis — May 2024
Focus: Reinforcement Learning, Deep Learning, Multi-Robot Systems, Optimal Control

D. J. Sanghvi College of Engineering — Mumbai, IN
Bachelor of Engineering, Mechanical — May 2022

TECHNICAL SKILLS

Languages	Python, Embedded C/C++, SQL, Bash, Terraform, Groovy
Software	Docker, ROS2, Jenkins, Git, Kubernetes
Frameworks	PyTorch, FastAPI, OpenCV, Tesseract OCR, llama.cpp
AWS Services	Lambda, API Gateway, S3, SQS, EC2, VPC, DynamoDB, Redshift, Glue, ElasticCache

WORK EXPERIENCE

Enterprise Technology — 06/23 - Present
ML Ops and AI Development Engineer — Tempe, AZ

- **LLM & AI Application Development:**
  - Engineered a Python library facilitating access to over 45 Large Language Models (LLMs), incorporating routing between multi-modal processing pipelines using AWS Lambda and API Gateway.
  - Deployed an LLM inference solution on AWS Lambda utilizing C binaries (llama.cpp) achieving 10 tokens/sec.
  - Developed a multi-threaded Python application for a Raspberry Pi-based interactive system, integrating text-to-speech, speech-to-text, facial recognition, and local LLM deployment.
  - Fine-tuned LLM endpoints and developed proxy servers for a Model-as-a-Service framework, enabling integration and hosting of various open-source models.
  - Implemented API rate limiting and caching using Lua scripting on AWS ElastiCache Redis with sliding window.
- **Cloud Infrastructure & DevOps:**
  - Leveraged Terraform to automate the provisioning and management of scalable, multi-stage AWS infrastructure.
  - Architected CI/CD pipelines using parallel EC2 docker builds on Jenkins and Kubernetes for building and deploying containers. This reduced the time taken to build 28 docker images from 30 minutes to 2 minutes.
  - Designed a micro-service architecture involving AWS Lambda, DynamoDB, S3, and API Gateways, enhancing system scalability and performance. Setup OAuth2 integrations architecture for google drive.
  - Implemented monitoring and alert systems using SNS, reducing system downtime and managing tracking.
  - Automated cross-cloud service provisioning across AWS and GCP, ensuring secure access control.
  - Setup datamart by integrating redshift and DynamoDB via zero ETL integrations.
  - Designed firecrawl (webcrawling) caching mechanism to reduce the number of API calls made and save cost.

Bio-Inspired Robotics, Technology and Healthcare Lab — 12/22 - 05/24
Graduate Student Researcher - Thesis — Tempe, AZ

- Automated the setup and execution of 180 friction analysis experiments using a custom-designed 3-axis apparatus with 6-axis load cell, reducing setup time by 40%.
- Created a PID control system integrated with ROS2 for the UR-16e robotic arm, enabling adaptive load-carrying tasks.
- Designed a ROS-based SpaceMouse controller, improving control capabilities and enabling real-time manipulation

PROJECTS

Multi-Modal OCR Document Organizer (Papertrail) | Python, OCR, AWS, OAuth

- Building a digital document organizer with fine grained access control, secure sharing and group access.
- Architecting a multi modal (any file type) OCR endpoint for accurate file to text conversion as a service.
- Google OAuth2 Integration for drive and picker API for seamless integration.

Audio projection embeddings model |Pytorch, Audio embeddings, RAG

- Developed an advanced audio-to-text embedding projection pipeline to enable Retrieval-Augmented Generation (RAG) workflows to support audio queries without transcription.
- Trained a projection embedding model that maps a vector of 512 dimension to 1536 dimension, utilizing CLAP model and processor and fine tuned on the librispeech_asr dataset.

Dexterous Manipulation with a Robotic Hand | Reinforcement Learning, Actor Critic, Python, Linux
- Advantage Weighted Actor Critic algorithm to enhance the performance of a 6-DoF robotic hand. Achieving up to a 20% improvement in dexterous manipulation success rates.

Multi Robot Search & Rescue | ROS2, RTAB, OpenCV
- Developed a decentralized quadcopter swarm with Potential-Field and Frontier Exploration for 3D mapping.
- Validated the swarm's ability to produce 100x100 grid maps in Gazebo, simultaneously avoiding local minima.

Statistical Analysis of Machine Learning Algorithms for Fraud Detection | Deep Learning, RNN, One-Hot Encoding
- Leveraged one-hot encoding, and preprocessing strategies like SMOTE in Python to optimize model training.
- Validated models via feature importance assessment and statistical evaluations, resulting in a 97.2% accuracy rate.

UAV Line Follower Drone | Simulink, Edge Detection
- Line Follower function for the Parrot Mambo Mini-Drone, identified specific HSV values of the track within 20 ms.
- Deployed via simulink, with 95% accuracy rate. Implemented edge detection to detect the edges of the track.

EXTRA-CURRICULAR

DJS KRONOS INDIA 03/19 - 05/21
Vice Captain Mumbai, IN
- Led software design and co-led mechanical development for a high-performance 4WD ATV using Simulink, achieving a 17% boost in operational efficiency, and was recognized with the award for 2nd Best 4WD design.
- Developed a data acquisition (DAQ) system using the GSM SIM900 module on a Raspberry Pi Zero, integrating with ThingSpeak for remote telemetry, enabling real-time data streaming and analysis.

HACKATHONS 03/24 - 10/24
Evaluator Tempe, AZ
- Evaluated various AI Innovation Hackathons and Spark Tank Challenges. Mentored and guided the participants so they could integrate AI capabilities in their solutions.

AI Acceleration 05/24 - Present
Colleague Tempe, AZ
- Managed workflow and assigned critical tasks to colleagues as well as approved architecture plans for multiple features and guided the team through the implementation.
- Oversaw majority of the development work and ensure code quality, consistency and productivity across all APIs.
- Involved in strategic as well as sprint planning tasks for more open communication between engineering and product for better clarity.
- Made sure every member had active involvement in feature development and gave them guidance to build their own plans according to the requirements.