Siddartha Kodaboina

669-649-2373 | siddartha.kodaboina@sjsu.edu | LinkedIn/siddartha-kodaboina | github.com/Siddartha-Kodaboina

Projects

Gen AI based Time and Space Complexity Analyser Chrome Extension Github — Live Demo

- Fine-tuned Llama 3.2 1b model using Unsloth quantization techniques with 159 manually curated time and space complexity puzzles stored in Parquet format, deploying the optimized model on EC2 for efficient inference
- Created a browser extension integrating the fine-tuned model, capable of analyzing time and space complexity across 100+ coding platforms, enhancing code optimality intution for students

AI Does Leetcode Github — Live Demo

- Architected a scalable Django application integrating 5 AI/ML services (GPT-4, MURF AI, Judge0) with AWS (Lambda, S3, DynamoDB), implementing event-driven test case generation and solution verification, processing 25+ test cases per question

HR Virtual Assistant Github — Live Demo

- Engineered a RAG-based HR recruitment system utilizing LLaMA 3.2 and GPT-4 for semantic resume parsing, implementing static chunking (200 tokens, 15-char overlap) and dual embedding strategy in Pinecone VectorDB, reducing candidate screening time by 5 hours and achieving 92% accuracy in role-candidate matching through Streamlit AI interface

Correct Code AI Github

- Built a system to enhance LLM performance on new frameworks using Retrieval Augmented Generation (RAG). Implemented LlamaParse to extract data from documentation, stored embeddings in MindsDB, and utilized LLaMA 3.2 3B for generating accurate code examples, improving framework adoption and developer productivity

Experience

Machine Learning Engineer Intern | San Jose State University September 2023 – Present

- **LLM Framework:** Wrote a Python framework to generate question-answer datasets for language model fine-tuning, utilizing LlamaParse for URL parsing and LLaMA 3.2 for intelligent Q&A generation, outputting CSV formats
- **SJSU AI Advisor Virtual Assistant:** Developed a college advising chatbot by fine-tuning LLaMA 3.2 model with Unsloth quantization, leveraging Python framework to scrape SJSU portals, improving students query response times by 48 hours

Software Engineer II (Backend Infrastructure Engineer) | Sandvine January 2022 – July 2023

- **Storage Monitoring:** Engineered a Python framework that monitors storage on build servers and purges old builds, reducing storage consumption by 30GB monthly
- **CI Log Intelligence:** Implemented comprehensive CI vulnerability detection framework with Python and Elasticsearch, identifying common failure patterns across pipelines, reducing MTTR (Mean Time to Resolution) by 40%
- **Parallel Processing:** Devised distributed mutex system in Python and Redis to streamline concurrent merge requests in cross-team repositories, orchestrating 1000+ monthly pipeline merges with zero conflicts
- **CI Optimization:** Analyzed the root cause for over 10 critical bugs in CI infrastructure, built the proof of concepts, and implemented the resolution by coordinating with stakeholders, these fixes increased the reliability of the CI infrastructure.

Software Engineer | Sandvine July 2021 – January 2022

- **CMS Platform:** Developed and deployed a large-scale documentation management system using React and Node.js, serving over 300K pages across 13 product lines, resulting in a 65% reduction in content management overhead
- **Performance Optimization:** Implemented redis caching resulted in decreasing page load time from 2.5s to 0.5s and reducing MySQL database query hits by 70%
- **Automation:** Engineered automated document processing including cleaning, editing, and publishing using Python, reducing processing time from 40 to 2 minutes and achieving 95% accuracy in content standardization
- **API Infrastructure:** Wrote Node.js/Express-based micro-services with 50+ RESTful APIs handling 20K+ daily requests, featuring role-based access control for diverse user groups and achieving 100ms average response time

Software Engineer Intern | Sandvine May 2021 – July 2021

- **Enterprise Distribution Platform:** Enhanced SpringBoot microservices architecture, delivering 20+ RESTful APIs handling 1000+ daily downloads across 500+ global customers, achieving 150ms average response time
- **Frontend:** Developed a responsive user interface with 15+ reusable components and advanced filtering system, boosting user engagement by 65% and garnering 98% positive feedback

Technical Skills

Programming Languages: Python, C++, Javascript, Typescript, Java
Frameworks: PyTorch, Tensorflow, Huggingface, Scikit-Learn, Pandas, Numpy, Django, Flask, FastAPI, Node, Spring Boot
Databases: Pinecone, ChromaDB, PostgresDB, MongoDB, DynamoDB, FireStore, GraphQL, Restful API, gRPC, Redis
Software Development Tools: Git, Github, Jenkins, AWS (EC2, Lambda, S3, Transcribe), Terraform, Kubernetes, Docker

Education

Master of Science, Computer Science Aug 2023 – Present
San Jose State University, United States of America

Bachelor of Technology, Computer Science Jul 2017 – Jul 2021
Jawaharlal Nehru Technological University, India